LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

numero 37
settembre 1971
anno IV
volume VII

La prospezione geochimica

La prospezione geochimica, con l'ausilio dell'analisi e della elaborazione automatica dei dati raccolti, integra i tradizionali criteri nella individuazione di mineralizzazioni economicamente sfruttabili

di Mario Dall'Aglio

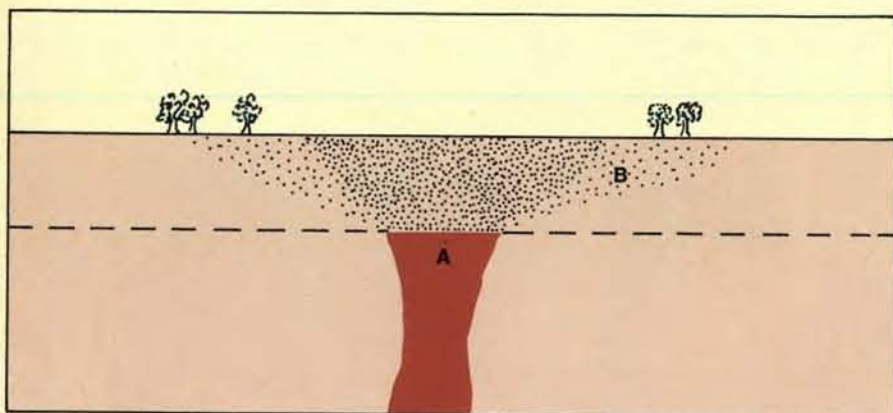
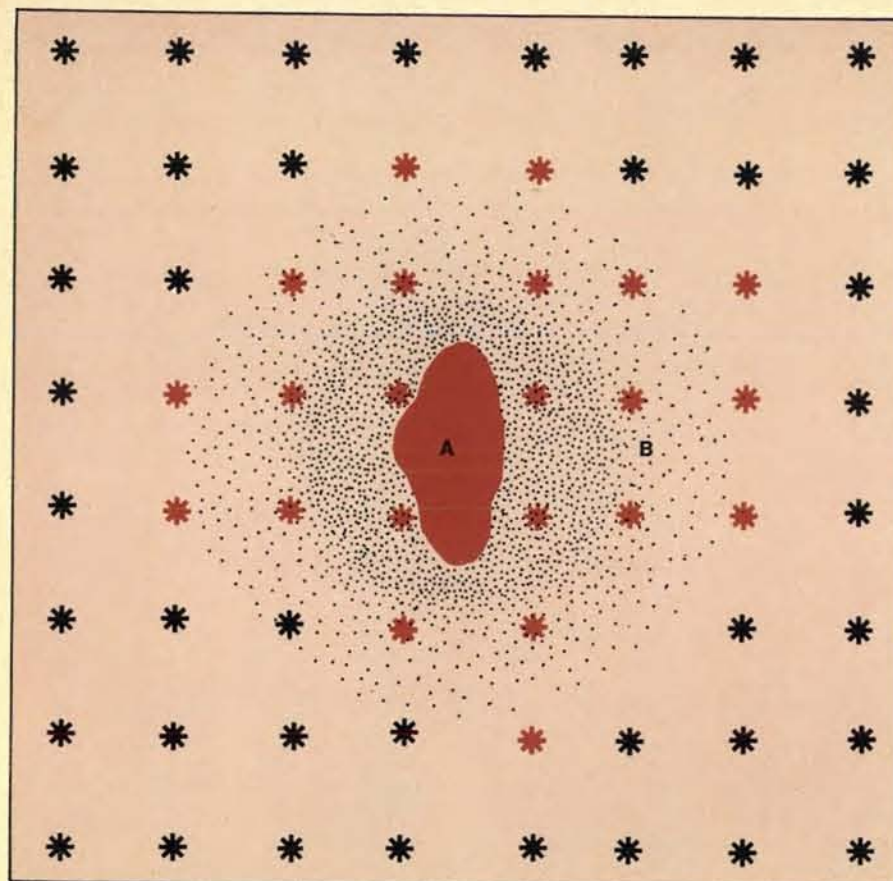
Nel corso dell'ultimo decennio si è verificato, nel mondo intero, un accrescimento eccezionale nel settore della ricerca mineraria. Dopo mezzo secolo di relativa stasi, gli stanziamenti e gli impegni assunti nel settore della ricerca mineraria hanno raggiunto livelli ragguardevoli tanto nei paesi industrializzati che in quelli in via di sviluppo. Questo aumentato interesse si riflette nell'estensione a zone sempre piú vaste dei metodi di prospezione già acquisiti e nei programmi di ricerca per la messa a punto di nuove tecniche di prospezione che rispondano ai requisiti di una maggiore efficienza.

Estese ricerche sono state condotte o intraprese tanto in aree relativamente nuove (Australia, Canada, Sud America, URSS, ecc.), quanto in aree già studiate in precedenza (Polonia, Francia, Stati Uniti, Inghilterra, ecc.).

Le cause di questo nuovo impulso nella ricerca mineraria sono numerose, interdipendenti e possono essere cosí elencate: 1) Per molte industrie la garanzia della continuità del flusso, della qualità e del prezzo negli approvvigionamenti delle materie prime riveste un ruolo primario ai fini di una razionale programmazione della produzione. Alcune industrie di primaria importanza, prime fra tutte quelle metallurgiche, tendono ad ampliare in senso verticale la loro produzione, sino a comprendere la fase dell'estrazione dei minerali. 2) L'industria estrattiva fornisce dei prodotti non ancora utilizzati, il cui valore non cessa di crescere nel corso dei successivi trattamenti che li trasformano in prodotti finiti. Il decollo dell'industria mineraria è suscettibile di innescare un piú ampio processo di industrializzazione. Per questo motivo le Nazioni Unite, l'UNESCO e altre organizzazioni internazionali dedicano sforzi ragguardevoli, sia tecnici che finanziari, per favorire la ricerca mineraria nei paesi in via di sviluppo. 3) Si è molto insistito sulla esiguità degli utili consentiti dall'industria mineraria. Per decenni tale industria è stata additata come una delle piú povere. In effetti le piú recenti esperienze hanno dimostrato che, con una conduzione moderna della strategia della ricerca e della coltivazione, si possono raggiungere utili elevati, superiori cioè al 10 % del fatturato. 4) Il mercato di molti minerali è caratterizzato da un deciso aumento dei fabbisogni, e quindi della richiesta, contrapposto al progressivo impoverimento delle riserve accertate e alla diminuzione delle scorte praticamente disponibili. Inoltre nel settore dell'industria mineraria i tempi che intercorrono tra l'inizio delle ricerche e la messa in funzione degli impianti estrattivi sono in genere lunghi, raggiungendo la decina d'anni. Risulta pertanto necessario programmare per tempo le ricerche di nuove fonti di approvvigionamento quando le scorte diminuiscono.

Alle osservazioni sopra riportate va aggiunto che tutti i paesi a intensa industrializzazione, con l'eccezione della sola Unione Sovietica, sono debitori verso l'estero della maggior parte delle materie prime minerali necessarie allo sviluppo della loro economia. L'attuazione di un vasto piano di ricerche geominerarie da parte di imprese private, statali e internazionali appare quindi la piú logica nell'attuale congiuntura del mercato dei minerali, solo che si consideri il fatto che la quasi totalità dei giacimenti attualmente sfruttati sono o quelli che affioravano, o quelli che presentavano manifestazioni in superficie.

Da seimila anni a questa parte la tecnica di ricerca mineraria piú seguita è stata la ricognizione in superficie. La maggior parte delle miniere in coltivazione sono state rinvenute seguendo gli stessi metodi di ricerca adottati dagli etruschi tre millenni orsono; i minerali vengono cioè estratti da corpi mineralizzati che presentano in superficie chiare manifestazioni della loro presenza. Non a caso nella provincia mineraria della Toscana la gran parte dei giacimenti ora in coltivazione, o sfruttati in tempi molto recenti, erano stati già scoperti dagli etruschi.

In effetti la gran parte dei giacimenti minerari esistenti nella crosta terrestre si sono formati nel sottosuolo. Di questi solo una piccola parte è stata portata alla luce da eventi orogenetici, o di sollevamento in genere, e dalla erosione, dopo la loro formazione. La potenzialità dei giacimenti non affioranti esistenti nella crosta terrestre è sicuramente molto grande e anche su questi si dovrà indirizzare l'industria estrattiva per soddisfare i fabbisogni futuri. Né bisogna credere che i giacimenti sepolti siano necessariamente antieconomici. Bisogna inoltre tenere presente che alcuni elementi possono raggiungere concentrazioni economicamente sfruttabili anche in assenza di

Rappresentazione schematica, in pianta (*sopra*) e in sezione (*sotto*), di come si può giungere alla individuazione di un corpo mineralizzato per mezzo della prospezione geochimica. Gli elementi concentrati in un filone mineralizzato (*A*) vengono dispersi dai processi esogeni. Questa azione, che si esplica soprattutto nella parte più vicina alla superficie, dà luogo a una aureola o alone di dispersione (*B*). L'esame dei campioni di terreno ha messo in luce la presenza di campioni anomali (*in colore*) circondati da campioni normali (*in nero*). L'analisi della distribuzione dei campioni anomali indica la presenza e l'estensione della aureola di dispersione e quindi della mineralizzazione.

manifestazioni osservabili a occhio. Tipico al riguardo il caso delle mineralizzazioni di oro, nelle quali il metallo risulta disperso nella ganga in particelle talmente minute (dell'ordine dei centesimi di micron) che non possono essere riconosciute che mediante l'analisi chimica (« oro invisibile »).

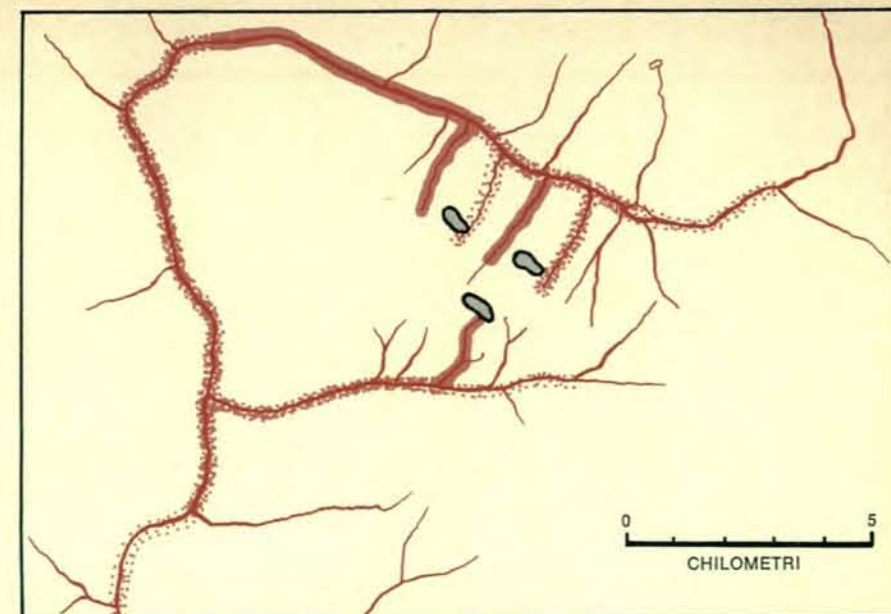
La ricerca dei giacimenti può essere impostata con due differenti criteri. In primo luogo per trovare nuovi giacimenti minerari risulta estremamente utile conoscere come, quando e dove essi si sono formati. Le osservazioni e le statistiche concernenti le mineralizzazioni conosciute mostrano alcune regolarità rimarchevoli nella presenza di giacimenti in formazioni geologiche corrispondenti a determinate epoche, o a determinate condizioni paleogeografiche e/o determinate situazioni stratigrafiche. Per analogia i nuovi giacimenti hanno una maggiore probabilità di essere rinvenuti in situazioni simili a quelle in cui sono stati riscontrati i giacimenti conosciuti. Anche se questo criterio ha condotto a risultati positivi, è bene sottolineare che criteri di questo tipo hanno costituito a volte una remora allo sviluppo scientifico dei metodi di ricerca e sono stati alla base di applicazioni sostanzialmente errate. Ciò discende dal fatto che i criteri operativi di ricerca sono stati elaborati sulla base dell'esperienza acquisita sui giacimenti noti e attraverso una impostazione modellistica dei processi minerogenetici. L'impostazione modellistica nelle scienze geologiche non risulta però sempre adeguata alla multiforme realtà naturale. Un modello minerogenetico rappresenta solo una parte dei caratteri comuni che emergono dall'esperienza acquisita, ma non rappresenta certamente l'insieme delle condizioni necessarie e sufficienti per la formazione di un giacimento minerario. Seguendo il secondo criterio, la ricerca di giacimenti può essere basata su metodologie che, anche prescindendo da ipotesi genetiche, giungono alla individuazione dei giacimenti stessi. Questi metodi di prospezione possono essere « indiretti » o « diretti »; sono chiamati indiretti quei metodi che ricorrono alla determinazione di alcune caratteristiche chimiche o fisiche degli elementi o dei composti ricercati; sono invece diretti i metodi che si servono della determinazione dei minerali o degli elementi ricercati in quanto tali.

Un giacimento minerario è rappresentato da una concentrazione eccezionale, rispetto al tenore medio delle rocce, di uno o più elementi in uno spazio limitato della crosta terrestre. I processi che portano alla concentrazione degli elementi nei giacimenti sono processi naturali che, in circostanze particolari, hanno portato alla separazione e alla concentrazione di alcuni elementi chimici, fino a costituire dei corpi mineralizzati anche imponenti, con centinaia di migliaia o addirittura milioni di tonnellate di metalli, contenuti in tenori anche centomila volte più elevati rispetto al contenuto medio delle rocce. Questi processi di concentrazione si possono verificare tanto nel ciclo interno quanto in quello esterno della materia; vale a dire sia durante i fenomeni di differenziazione magmatica, di metamorfismo e di circolazione termale (giacimenti endogeni), sia nel corso dei processi di erosione, trasporto e sedimentazione (giacimenti esogeni).

Come si è già accennato, sino a tempi relativamente recenti, la ricerca mineraria è stata condotta con criteri prevalentemente geologici e ricorrendo quasi esclusivamente all'indagine di superficie. In particolare i giacimenti minerari sono stati ricercati soprattutto in condizioni geologiche e ambientali simili a quelle delle già note miniere in coltivazione. D'altro canto le miniere note erano state aperte in corrispondenza delle manifestazioni superficiali più vistose. La ricerca indirizzata prevalentemente su particolari formazioni geologiche (o tipi di rocce) poteva quindi portare al rinvenimento di mineralizzazioni solo in tali formazioni.

A parte la considerazione preliminare che non sempre le manifestazioni più vistose in superficie corrispondono alle mineralizzazioni più importanti, è necessario aggiungere che la scarsa conoscenza in termini quantitativi dei processi geochimici di trasporto e di accumulo degli elementi ha portato per il passato ad attribuire prevalentemente ai processi magmatici o termali – connessi cioè al movimento di rocce fuse e alla circolazione di soluzioni ad alta temperatura a esse connesse – la causa della formazione dei corpi mineralizzati. In altri termini i processi magmatici e termali sono stati spesso invocati come dei « deus ex machina » capaci di mascherare una sostanziale ignoranza dei termini quantitativi dei processi minerogenetici. Le ricerche geochimiche hanno invece dimostrato che molti elementi potevano circolare nell'ambiente superficiale in concentrazioni molto basse in assoluto, ma sufficienti per originare, in lunghi periodi di tempo, arricchimenti di importanza economica.

Quanto sopra affermato trova corrispondenza nella storia, per esempio, delle ricerche di minerali uraniferi. Sino al 1955 le ricerche sono state indirizzate quasi esclusivamente sulle formazioni granitiche e sulle pegmatiti perché le prime miniere d'uranio importanti erano state aperte in tali formazioni. Col progredire degli studi geochimici e minerogenetici si accertò che l'uranio poteva circolare nelle acque naturali in quantità considerevoli e poteva precipitare selettivamente in corrispondenza di alcuni orizzonti riducenti dando luogo a concentrazioni minerarie economicamente importanti. Oggi la maggior parte dei giacimenti uraniferi sfruttati è di origine sedimentaria.

Esaminando il problema in termini più generali, un obiettivo importante delle ricerche geochimiche è costituito dalla comprensione e descrizione completa delle leggi che regolano la distribuzione e la circolazione degli elementi nella crosta terrestre. La comprensione di tali processi di trasporto e di accumulo degli elementi costituisce la base logica e immediata per ricostruire i processi che hanno portato alla formazione dei giacimenti minerari, e per indirizzare quindi le ricerche pratiche dei diversi minerali utili.

Grande importanza ha nella prospezione geochimica lo studio delle scie di dispersione che si formano nelle acque superficiali e nei depositi fluviali in corrispondenza delle mineralizzazioni. Nell'esempio raffigurato, l'anomalia più intensa (*colore pieno*) deriva direttamente dalle mineralizzazioni (*macchie grige*) che danno origine a valori di anomalia via via decrescenti con l'allontanarsi dalle zone mineralizzate (*puntinato in colore*).

Un altro aspetto interessante è costituito dalla circostanza che molti elementi presentano caratteristiche chimiche simili tra di loro. A seguito dei processi di combinazione e di trasporto, continuamente in atto in quel grande laboratorio geochimico che è l'ambiente naturale, tali elementi circolano con le stesse modalità, e sono quindi suscettibili di migrare insieme e di concentrarsi in zone coincidenti o poco distanti l'una dall'altra. Lo studio delle associazioni o « paragenesi » degli elementi nei diversi ambienti naturali ha da un lato permesso di chiarire meglio le leggi che regolano il loro trasporto in natura, e dall'altro di prevedere la presenza di mineralizzazioni, anche importanti, laddove sussistevano indicazioni di arricchimenti di elementi geochimicamente associati.

Sotto questo profilo i ricercatori dell'Unione Sovietica hanno condotto le ricerche più estese al fine di fornire un quadro unitario dei processi geochimici che regolano la circolazione degli elementi nella crosta terrestre, arrivando alla composizione di un quadro unitario analogo a quello che Mendeleev aveva costruito nel 1869 per gli elementi chimici. L'accostamento con il grande chimico russo del XIX secolo non è casuale perché molte osservazioni di tipo geochimico sono alla base della tabella di Mendeleev e perché le similitudini e le differenze nelle caratteristiche chimiche degli elementi, sintetizzate nella tabella di Mendeleev, costituiscono il punto di partenza per lo sviluppo delle conoscenze geochimiche.

I criteri geochimici costituiscono dunque la base per la comprensione dei processi di accumulo che hanno portato alla formazione dei giacimenti minerari, e per guidare le ricerche pratiche col prevedere quali sono le condizioni e gli ambienti naturali che presentano le probabilità maggiori di contenere i giacimenti dei diversi elementi.

Le conoscenze e i criteri geochimici hanno una funzione importante anche nella fase più strettamente applicativa della ricerca dei giacimenti. Prospezione geochimica è stata chiamata questa branca applicata della geochimica.

È stato già ricordato che un giacimento minerario rappresenta una concentrazione, spesso eccezionale, rispetto al tenore medio della crosta terrestre di uno o più elementi in uno spazio limitato. Un giacimento è comunque sempre sottoposto, in maniera più o meno marcata, a fattori che tendono a disperdere gli elementi concentrati nel giacimento stesso. Anche durante gli eventi minerogenetici che portano alla messa in posto dei corpi mineralizzati, una parte degli elementi si disperde attorno al giacimento dando luogo agli « aloni primari ». In special modo i giacimenti che si vengono a trovare nella parte più superficiale della crosta

TIPO DI PROSPEZIONE	SCOPO DELLA RICERCA	SOSTANZE CAMPIONATE	MAGLIA DI CAMPIONAMENTO	NUMERO DI CAMPIONI RACCOLTI PER 100 km²
RICOGNIZIONE REGIONALE	Ricognizione preliminare di vaste aree poco note	Acque superficiali e depositi alluvionali	10-5 km	1-10
PROSPEZIONE STRATEGICA	Individuazione, in zone già indiziate, di aree anomale promettenti	Acque superficiali, depositi alluvionali	1 km	50-200
PROSPEZIONE TATTICA	Scoperta e delimitazione di mineralizzazioni	Suoli, acque di sorgenti, gas del sottosuolo, rocce	100-10 m	10 000-500 000

In questa tabella sono riportate le caratteristiche essenziali dei piú noti metodi di cui si avvale la prospezione geochimica. I tre tipi riportati rappresentano le fasi classiche nelle quali si articola la ricerca di nuove zone atte a uno sfruttamento minerario.

terrestre sono sottoposti a una intensa azione di dispersione degli elementi da parte dell'erosione, dell'alterazione e lisciviazione da parte delle acque naturali. Tutti questi processi tendono a disperdere gli elementi contenuti nelle mineralizzazioni, in un'ampia zona attorno al giacimento stesso. Nelle vicinanze dei giacimenti si vengono quindi a formare le « aureole » o « aloni di dispersione ». Nell'ambito di una aureola le concentrazioni degli elementi diminuiscono gradatamente man mano che ci si allontana dalla mineralizzazione, sino a raggiungere i tenori normali della regione considerata (*si vedano le figure alle pagine 14 e 15*).

Gli aloni di dispersione hanno estensione piú vasta del giacimento da cui traggono origine. La prospezione geochimica si basa proprio su questo semplice principio: l'analisi di campioni di sostanze naturali, come acque, suoli, piante e rocce, permette di individuare l'esistenza di un alone di dispersione e quindi la probabile presenza della mineralizzazione da cui l'aureola ha preso origine.

Lo schema sopra proposto è volutamente semplificato, al fine di illustrare il principio su cui si basa la prospezione geochimica dei giacimenti. In realtà, quando il geochimico si trova a dover interpretare i dati analitici provenienti da una campagna di prospezione, le cose si presentano in maniera molto piú complessa. Basti pensare alla circostanza che nei mezzi correntemente campionati nell'ambito dei rilievi geochimici interagiscono, a volte in maniera determinante, processi legati all'attività biologica della flora e della fauna, e alla reattività dei loro prodotti di decadimento. Cosí alcuni elementi presenti nelle acque possono essere catturati selettivamente dai microorganismi e dalle sostanze organiche provenienti dalla decomposizione della materia vivente; in questa maniera può risultare alterata la regolare formazione delle aureole di dispersione. Esistono inoltre processi e fenomeni naturali capaci di mascherare l'esistenza, e soprattutto l'individuazione, delle aureole di dispersione. Basterà citare le variazioni litologiche, idrologiche e climatiche. Persino l'inquinamento dell'ambiente dovuto all'attività umana provoca spesso intense anomalie nelle concentrazioni degli elementi ricercati (*si veda la figura nella pagina a fronte*).

I metodi di prospezione geochimica presentano una vasta gamma di applicazioni e possono soddisfare una notevole varietà di esigenze programmatiche: dalle ricerche preliminari su aree poco note e molto vaste, alle ricerche di grande dettaglio sulle mineralizzazioni affioranti. È invalso l'uso di suddividere i metodi di prospezione in base al tipo di materiale naturale esaminato, e quindi in base al tipo di aureola di dispersione studiata: cosí viene chiamata idrogeochimica la prospezione basata sull'analisi di campioni di acque, pedogeochimica quella basata sull'analisi di campioni di suoli, prospezione alluvionale quella basata sull'analisi di campioni di depositi fluviali, prospezione geobotanica quella che ricorre all'analisi di parti di piante, ecc.

Al fine di esaminare le caratteristiche dei diversi metodi di prospezione geochimica è però piú utile prendere in considerazione allo stesso tempo lo scopo che si prefigge la ricerca, i risultati che possono essere raggiunti, il costo dell'operazione e la maglia di campionamento adottata. Esiste naturalmente interdipendenza tra la maglia di campionamento, il mezzo geochimico campionato, le dimensioni dell'aureola, e quindi i risultati che possono essere conseguiti. L'estensione dell'alone di dispersione e le sue caratteristiche sono infatti fattori molto importanti ai fini dell'economia dei lavori di prospezione geochimica. Cosí aloni di dispersione molto estesi possono essere individuati adottando una maglia di campionamento proporzionalmente ampia. D'altro canto sia l'estensione sia la forma degli aloni di dispersione dipendono, oltre che dalle caratteristiche delle mineralizzazioni e dalle proprietà geochimiche dell'elemento cercato, anche dai differenti mezzi che vengono scelti per il campionamento, e dalle condizioni climatiche e morfologiche. Cosí, per esempio, gli aloni idrogeochimici degli elementi mobili sono gli aloni piú estesi e piú regolari; ma in climi aridi gli aloni idrogeochimici non si sviluppano regolarmente e i depositi alluvionali possono sostituire – come mezzo di campionamento – le acque superficiali. Per meglio classificare i diversi metodi di prospezione geochimica è opportuno quindi prendere in considerazione, da un lato, i diversi mezzi geochimici secondo l'ordine di grandezza delle aureole di dispersione – e della maglia piú ampia quindi che può essere adottata nel campionamento con la sicurezza di rinvenire anomalie in corrispondenza delle aureole di dispersione – e dall'altro tener conto dello scopo che viene prefisso alla ricerca. La dimensione degli aloni di dispersione è dell'ordine delle decine di chilometri per l'ambiente idrogeochimico superficiale, della decina di chilometri per i depositi fluviali, e delle centinaia di metri nei suoli. Le cifre riportate rappresentano solo degli ordini di grandezza; col variare delle condizioni am-

Nella cartina della pagina a fronte sono riportati in forma sintetica i risultati ottenuti nel corso di un rilievo geochimico a maglia regionale, basato sulla determinazione del mercurio nei depositi fluviali, effettuato sull'intera area della Toscana nel 1966. Accanto alle aureole di dispersione rinvenute in corrispondenza di giacimenti di cinabro in corso di sfruttamento o di mineralizzazioni comunque note (*anomalie A e B*), va messo in evidenza che alle anomalie *D* e parzialmente *C*, non corrispondevano segnalazioni minerarie. Anche in regioni notoriamente mercurifere e con antica tradizione mineraria, i metodi diretti di ricerca, e in primo luogo la geochimica, possono aprire nuovi orizzonti alle prospettive minerarie. È infine da segnalare l'anomalia *E* che non corrisponde alla presenza di alcuna mineralizzazione ma è causata da scarichi industriali e artigianali contenenti $HgCl_2$. I depositi fluviali sottraggono il mercurio dalle acque naturali assumendo quindi tenori anomali. Questo esempio mostra come i risultati conseguiti dalle prospezioni geochimiche possono fornire un valido contributo non solo alla scoperta di eventuali giacimenti, ma anche per il controllo degli inquinamenti.

Alcune attrezzature analitiche impiegate per le prospezioni geochimiche. Il piccolo laboratorio portatile (*sopra*) che trova alloggio in un contenitore di alluminio di limitate dimensioni e peso (circa 2 kg) è impiegato per l'analisi in campagna dei metalli pesanti (Cu, Zn e Pb), delle acque naturali, dei suoli, dei depositi alluvionali e delle rocce. Con questa semplice attrezzatura, del costo di poche decine di migliaia di lire, possono essere effettuate una trentina di analisi al giorno con precisione e selettività non molto elevate. Nella foto sotto è raffigurata un'altra attrezzatura per la determinazione di Zn e Cu nei campioni di sostanze naturali. Con questa strumentazione, del costo di qualche centinaio di migliaia di lire, è possibile effettuare una quarantina di analisi al giorno con risultati di buona qualità. Questa attrezzatura, facilmente smontabile e trasportabile, può essere utilizzata, all'occorrenza, anche in laboratori di fortuna.

bientali e a seconda degli elementi ricercati si possono avere divergenze anche notevoli da questi valori. Per quanto concerne lo scopo da raggiungere, i rilievi geochimici possono essere attuati tanto per addivenire a un inventario preliminare sulla convenienza o meno di eseguire nell'area in oggetto ricerche di maggior dettaglio, quanto per delimitare una zona che riserva alte probabilità di contenere un giacimento minerario, quanto ancora per giungere alla scoperta e alla delimitazione di mineralizzazioni.

In base a quanto esposto in precedenza tre sono i metodi di prospezione geochimica che sono generalmente adottati: la ricognizione regionale, la prospezione strategica e la prospezione tattica (*si veda la tabella riportata a pagina 16*).

Con il metodo della ricognizione regionale si può procedere alla ricerca, su scala regionale, dell'esistenza o meno di zone promettenti dal punto di vista minerario. Lo scopo può essere anche un giudizio preliminare sulla convenienza o meno di investire ulteriori fondi per ricerche di maggior dettaglio, o stabilire infine nella regione nella quale è stata fatta la prospezione un ordine di precedenza tra le zone che presentano probabilità diverse di successo. Per ottenere questo scopo la prospezione regionale si rivolge allo studio degli aloni di dispersione piú estesi e persistenti, e quindi all'esame delle acque superficiali e ai depositi fluviali. La maglia di campionamento è di circa 5-10 km sul reticolo idrografico, il che corrisponde – in condizioni medie – a un campione ogni 20-100 km^2. Le zone indiziate che vengono indicate da tale tipo di prospezione sono estese, dell'ordine cioè di molti chilometri quadrati.

Lo scopo che ci si prefigge con la prospezione strategica consiste nella delimitazione di zone dell'ordine del chilometro quadrato nelle quali è molto probabile rinvenire mineralizzazioni. Anche in questa fase le aureole di dispersione piú ricercate sono le idrogeochimiche e le alluvionali. La maglia di campionamento è di circa un chilometro sul reticolo idrografico, e la densità di campionamento risulta di un campione ogni 0,3-3 km^2. La prospezione strategica viene applicata in zone già indiziate o da motivazioni geologiche e minerogenetiche, o da indizi raccolti nella fase di ricognizione regionale.

I risultati che vengono conseguiti con la prospezione tattica giungono in questo caso sino alla scoperta dei giacimenti affioranti o alla ubicazione dei sondaggi di accertamento minerario. La maglia di campionamento può variare da centinaia di metri a pochi metri; di conseguenza la densità di campionamento varia da 100 a piú di 10 000 campioni per chilometro quadrato. In questa fase vengono studiati prevalentemente i suoli, le rocce, le acque freatiche e i gas del sottosuolo.

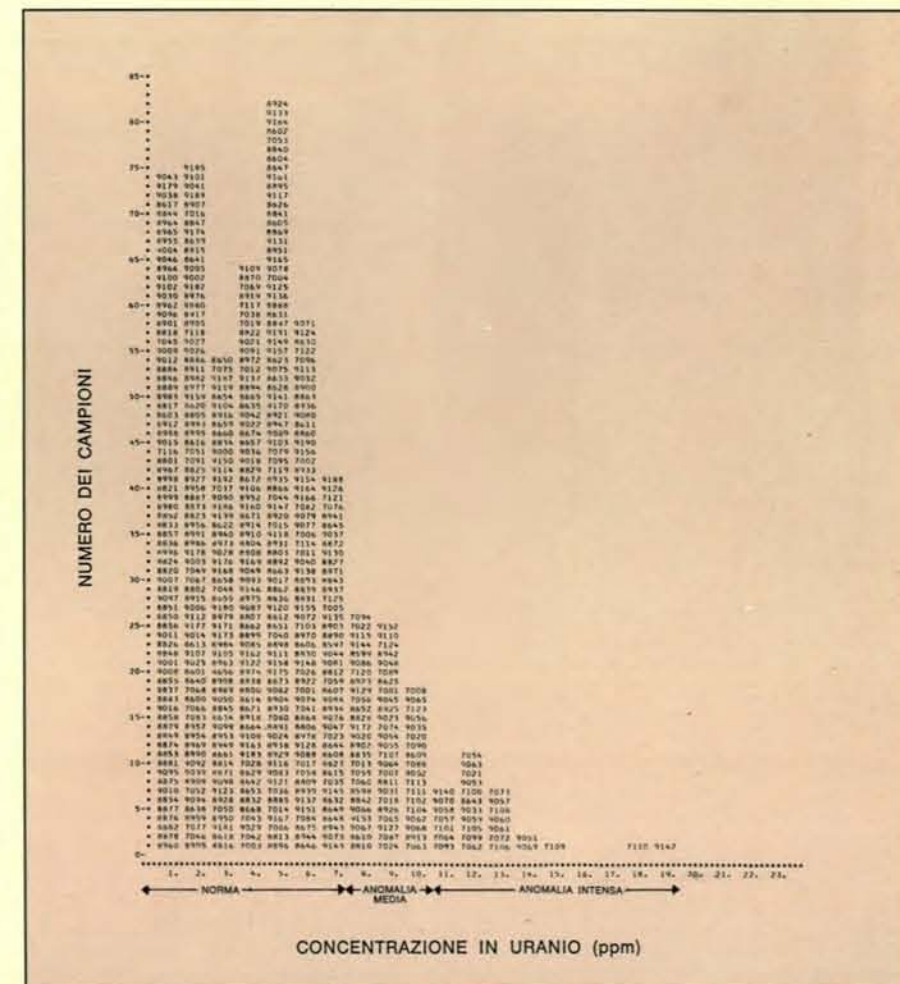
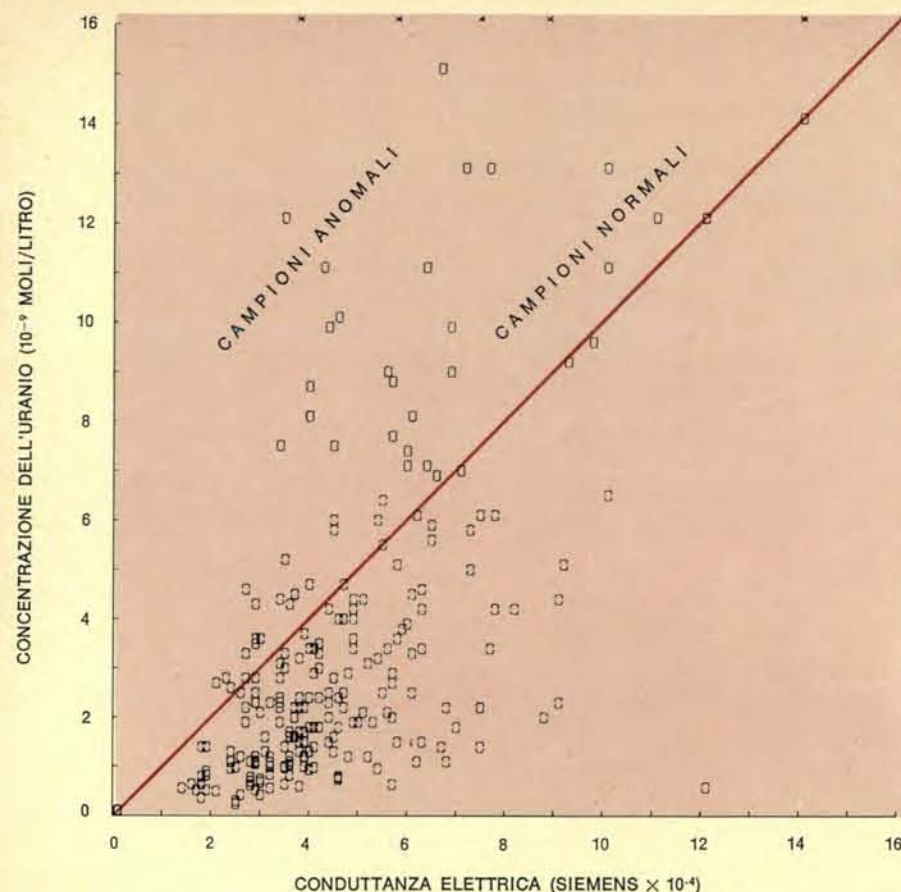
Anche se esistono dei termini di passaggio tra le tre fasi sopra menzionate – termini che coprono con continuità tutta la gamma compresa tra i limiti estremi – ci si riferisce in genere a questi tre tipi di prospezione perché ciascuno di essi rappresenta uno dei momenti programmatici nei quali si articola la ricerca mineraria.

Pur senza entrare in un esame dettagliato della programmazione dei diversi stadi del lavoro e della conduzione dei rilievi, è utile soffermarsi brevemente su alcune fasi del lavoro che presentano esigenze e condizioni operative peculiari di questo genere di studi.

La raccolta dei campioni è la prima fase operativa dei rilievi geochimici. La corretta conduzione del lavoro sul terreno è essenziale per non compromettere in maniera irreversibile la validità delle conclusioni che vengono tratte al termine della campagna. È necessario in particolare che i campioni raccolti siano rappresentativi del mezzo geochimico in studio e che non intervengano fattori di perturbazione capaci di alterare gravemente la fenomenologia delle aureole di dispersione. Nel corso di rilievi basati sull'esame di campioni di suoli residui è necessario, per esempio, campionare sempre lo stesso orizzonte pedologico. Dato che i diversi strati del terreno (orizzonti) hanno profondità diverse a seconda delle condizioni locali, nel corso di una stessa campagna è possibile raccogliere inavvertitamente campioni corrispondenti a orizzonti pedologici diversi. In tal maniera l'analisi dei prelievi può mettere in luce anomalie nel tenore dell'elemento ricercato, che non dipendono affatto dalla presenza di una aureola di dispersione, ma solo da una erronea variazione dell'orizzonte campionato.

Se si considera infine che la fase di campionamento viene sovente effettuata in zone disagiate, in cui risulta molto costoso ritornare per effettuare controlli o ripetere anche parzialmente i prelievi, si comprende l'importanza che riveste la fase del prelievo dei campioni al fine della buona riuscita della campagna.

Nell'ambito dei rilievi geochimici un posto preminente occupa l'analisi chimica delle sostanze naturali. Gli elementi da analizzare sono presenti in concentrazioni molto basse, anche inferiori a

Complesso spettrofotometrico ad assorbimento atomico ad alto livello di automazione utilizzato per l'analisi di Pb, Zn, Cu, e di molti altri elementi. Con tale attrezzatura, del costo di una ventina di milioni, è possibile effettuare circa 300 analisi al giorno con ottime caratteristiche di precisione e selettività. Con questo complesso l'intervento manuale è molto ridotto: una serie di 100 campioni viene prima portata alla diluizione voluta e poi sottoposta all'analisi in maniera del tutto automatica. I segnali in uscita possono essere registrati graficamente, numericamente o su nastro perforato. Il nastro può essere letto da un calcolatore che provvede alla elaborazione dei dati.

In questo istogramma relativo ai tenori in uranio dei depositi fluviali di una intera regione, costruito e stampato direttamente da un elaboratore elettronico, sono indicati gli intervalli di tenore corrispondenti ai diversi gradi di anomalia riscontrata nei campioni. La selezione dei campioni anomali, cioè di quei campioni raccolti nelle aureole di dispersione, è la fase piú importante durante lo svolgimento dei rilievi geochimici.

In questa figura e in quella della pagina a fronte sono rappresentati sotto forma di diagrammi, disegnati direttamente da un elaboratore elettronico, i criteri piú completi per la selezione dei campioni anomali. I punti rappresentativi di campioni di acque naturali sono riportati in un diagramma uranio-conduttanza elettrica. I due parametri risultano correlati perché le acque si arricchiscono contemporaneamente tanto di uranio che di altri sali. I valori normali di uranio spaziano in un ampio intervallo di concentrazione e sono rappresentati nel diagramma dalla nube di punti lungo la diagonale. I campioni anomali sono quelli che si discostano in maniera significativa da tale nube.

una parte per milione nelle sostanze solide e a un microgrammo per litro nelle acque naturali. Una prima caratteristica dei metodi analitici che vengono impiegati nella prospezione geochimica è quindi quella di dover essere molto sensibili. Le sostanze da analizzare sono inoltre a matrice molto variabile e con contenuti apprezzabili di elementi che possono interferire nella determinazione dell'elemento ricercato. Le metodologie analitiche debbono quindi rispondere anche al requisito di una elevata specificità. A volte è necessario o opportuno che alcune analisi vengano eseguite sul posto stesso del prelievo, o poco tempo dopo la raccolta dei campioni. Questo sia per evitare che intervengano alterazioni dei campioni capaci di inficiare la significatività delle analisi, sia per poter disporre entro breve tempo dei risultati analitici e poterne quindi tenere conto per la migliore programmazione del lavoro ancora da svolgere. Una terza caratteristica delle metodologie analitiche è quella di poter essere eseguite con strumentazioni portatili che siano in grado di funzionare autonomamente anche a bordo di autovetture, o in una tenda, e persino in luoghi privi di qualsiasi servizio.

Nel corso dei rilievi geochimici è necessario eseguire un numero impressionante di analisi chimiche. Si pensi che nella sola Unione Sovietica vengono raccolti e analizzati ogni anno oltre 10 milioni di campioni di sostanze naturali, e che nell'ambito di una sola campagna della durata di pochi mesi risulta necessario eseguire da 10 000 a 100 000 analisi chimiche. I metodi analitici applicati alla prospezione geochimica debbono essere pertanto facilmente applicabili ai lavori di routine, permettendo un'alta velocità di lavoro anche in condizioni logistiche disagiate. Poiché è difficile reperire in commercio una strumentazione che risponda allo stesso tempo a tutti i requisiti sopra elencati è spesso necessario adattare le metodologie e persino modificare o costruire ex novo le apparecchiature analitiche necessarie. D'altro canto la disponibilità di metodologie e di strumentazioni analitiche ad hoc influisce in maniera determinante sulla validità dei risultati conseguibili mediante la prospezione geochimica e sul contenimento dei costi (*si vedano le figure a pagina 18 e in alto a pagina 19*).

Una volta che è stata completata la raccolta delle informazioni di campagna e di laboratorio, inizia la parte piú delicata dei rilievi geochimici, e cioè l'interpretazione di tutti i dati al fine di pervenire a delle valide conclusioni operative finali.

Nella fase interpretativa dei dati risulta particolarmente importante e critica la individuazione dei campioni « anomali », di quei campioni cioè che corrispondono a una aureola di dispersione e che indicano quindi la presenza di una mineralizzazione. Se si considera il tenore di un elemento in una serie di campioni si può osservare che i valori delle concentrazioni si presentano in genere con frequenze diverse. Ci sono alcuni valori che si presentano un maggior numero di volte, altri un minor numero, e altri ancora molto raramente. Nel linguaggio statistico la « norma » indica il valore che si presenta con maggior frequenza, e quindi che ha piú elevata probabilità di comparire. Nel caso si consideri un solo parametro, la definizione della norma risulta semplice e di conseguenza è semplice la selezione dei campioni anomali, di quelli cioè che si discostano in maniera significativa dalla norma stessa. Nella figura in basso a pagina 19 è riportato come esempio la distribuzione di frequenza, relativa al tenore di uranio in campioni di depositi fluviali, costruita direttamente da un elaboratore elettronico, con l'indicazione dei campioni anomali. In genere però la conoscenza del contenuto di un elemento chimico nelle sostanze naturali ha scarsa importanza se considerata da sola. Il tenore dell'elemento ricercato acquista un significato completo solo se considerato in relazione al tenore presentato dagli altri elementi nello stesso campione e nel contesto delle caratteristiche gografiche, geologiche e climatiche della zona in cui è stato raccolto il campione. In questo caso la definizione della norma, e la conseguente selezione dei campioni anomali, deve essere effettuata in maniera piú completa e articolata che non considerando i valori di un solo parametro. I criteri che vengono adottati si rifanno ai concetti di normale e anomalo cosí come vengono recepiti nel linguaggio comune. La definizione della normalità e dell'anomalia di un campione di sostanza naturale non è dissimile da quella normalmente adottata per l'aspetto fisico di un uomo. Il concetto di normale non è infatti legato alla sola altezza o alla sola larghezza o al solo peso di una persona, ma al verificarsi o meno di determinate proporzioni tra i diversi parametri fisici. Le figure riportate qui accanto e nella pagina a fronte illustrano meglio, col ricorso ad alcuni esempi concreti, le procedure e i criteri adottati nella selezione dei campioni anomali raccolti nel corso dei rilievi geochimici.

Una volta che è stata effettuata la selezione dei campioni anomali, la fase finale interpretativa viene completata dallo studio della distribuzione nell'area in oggetto dei campioni anomali in relazione alle condizioni geologiche, e alla luce delle caratteristiche geochimiche degli elementi ricercati. Le anomalie significative che vengono cosí individuate indicano infine la presenza di aureole di dispersione e quindi di mineralizzazioni.

È interessante sottolineare che i criteri geochimici applicati alla ricerca mineraria hanno costituito – e piú costituiranno per l'avvenire – un contributo essenziale all'impostazione unitaria e globale dello studio dei processi legati alla formazione dei giacimenti minerari e alla loro dispersione. L'insieme di tutti questi complessi processi, che di norma si svolgono nell'arco dei tempi lunghi della geologia, è, per sua natura, riconducibile sempre alle modalità e alle caratteristiche di circolazione degli elementi chimici sulla Terra.

In questo diagramma, fornito direttamente da un elaboratore come quello riportato nella pagina a fronte, è riportato il risultato del calcolo di correlazioni multiple effettuate numericamente ricorrendo alle tecniche dell'analisi di regressione. Il tenore in uranio è stato studiato in rapporto alla concentrazione di tutti i maggiori costituenti disciolti nei campioni di acqua (Ca^{++}, Mg^{++}, Na^{+}, K^{+}, HCO_3^{-}, SO_4^{--} e Cl^{-}). In ordinate sono riportati i valori dei tenori in uranio determinati sperimentalmente e in ascisse quelli calcolati in base all'analisi della regressione. I tenori anomali sono anche qui quelli che si discostano in maniera significativa dalla nube lungo la diagonale.

Alla luce di queste considerazioni si comprende anche come la suddivisione rigida tra ricerca applicata e di base risulta in questo campo spesso artificiosa, e conduce sovente alla cattiva utilizzazione delle risorse devolute a due settori che a torto vengono considerati come stagni e indipendenti, e che si ha interesse invece a far procedere insieme e con un continuo scambio di informazione al fine di un piú rapido ed economico progresso di entrambi i settori.

Nell'ambito delle campagne di prospezione geochimica vengono raccolti un numero impressionante di dati concernenti la distribuzione e la circolazione degli elementi chimici in natura. Questi dati portano a un approfondimento rimarchevole nella conoscenza della composizione chimica dell'ambiente naturale. Ciò è valido sia per quanto attiene al controllo degli inquinamenti naturali e artificiali, sia per quel che riguarda la distribuzione degli elementi in tracce (distribuzione che è importante conoscere anche per la programmazione della migliore utilizzazione agricola dei vari tipi di terreni). Non va dimenticato che in Unione Sovietica, attorno al 1935, la prospezione geochimica ha ricevuto un notevole impulso dalla necessità di conoscere la distribuzione di molti elementi in tracce ai fini di una corretta utilizzazione agricola dei diversi suoli.

Per completare il panorama del ruolo svolto dalla geochimica nell'ambito della ricerca mineraria, è utile dare uno sguardo ai rapporti tra ricerca mineraria e industria estrattiva che sono ragionevolmente prevedibili in base alle attuali tendenze.

È stato affermato che il futuro fabbisogno mondiale di minerali sarà cosí elevato che tutte le mineralizzazioni importanti saranno in breve esaurite e che l'umanità si approvvigionerà delle materie prime minerali partendo dalle rocce ordinarie ed estraendo tutti gli elementi in esse contenuti. L'industria mineraria sarà allora completamente assorbita dall'industria chimica. Questa affermazione avveniristica contiene qualcosa di vero. Le miniere, come sono state concepite per il passato, andranno probabilmente scomparendo. La tecnica mineraria a « cielo aperto », con l'ausilio di una buona meccanizzazione, ha dimostrato che è possibile estrarre materie prime minerali a minor costo pur partendo da minerali a tenori piú bassi di quelli considerati economici per il passato. Resta comunque il fatto che in natura sono riscontrabili arricchimenti di elementi chimici che raggiungono con facilità fattori maggiori di 10 e che sono estesi per intere province geochimiche. Un'industria mineraria, molto piú progredita dell'attuale, sarà guidata nelle sue decisioni sempre in maggior misura dall'accurato studio della distribuzione degli elementi nelle diverse parti della crosta terrestre. Questo perché i costi degli impianti e delle infrastrutture tenderanno ad aumentare considerevolmente; di conseguenza le decisioni relative all'attuazione di nuovi impianti dovranno tenere sempre in maggior conto le informazioni geochimiche.

In questa figura sono riportati (*in nero*) i movimenti oculari di un soggetto che guarda per la prima volta un disegno ricavato dal « Vecchio che pensa » di Paul Klee. I numeri mostrano l'ordine delle fissazioni effettuate dal soggetto sulla figura durante una parte dei 20 secondi di osservazione. Le linee tra i punti rappresentano i rapidi movimenti degli occhi da un punto di fissazione a quello successivo. I movimenti occupano circa il 10 per cento del tempo totale necessario per la visione.

I movimenti degli occhi e la percezione visiva

Le registrazioni dei punti sui quali si soffermano gli occhi nel guardare e del tracciato che essi seguono nell'osservare forniscono la chiave per la comprensione del processo con il quale il cervello riconosce un oggetto

di David Noton e Lawrence Stark

Gli occhi sono gli organi di senso piú attivi dell'uomo. Altri recettori sensoriali, come per esempio le orecchie, ricevono passivamente tutti i segnali in arrivo, mentre gli occhi si muovono di continuo per la scansione e l'ispezione dei particolari del mondo circostante. I movimenti dell'occhio hanno un ruolo molto importante nella percezione visiva, e la loro analisi può rivelare molte cose sulla natura del processo percettivo.

Abbiamo recentemente registrato i movimenti oculari di soggetti che dovevano prima osservare e poi riconoscere alcuni oggetti a loro sconosciuti. In sostanza abbiamo rilevato che ogni individuo ha un proprio modo caratteristico di guardare un oggetto familiare. I suoi occhi tendono a seguire un determinato tracciato per ogni oggetto che egli osserva o riconosce. I nostri risultati suggeriscono una nuova ipotesi sull'apprendimento visivo e il riconoscimento. Prima di descrivere e illustrare le nostre esperienze piú dettagliatamente, daremo un quadro generale dell'argomento riferendo alcuni esperimenti precedenti che hanno contribuito alla interpretazione dei nostri risultati.

I movimenti oculari sono necessari per una ragione fisiologica: una informazione visiva dettagliata può essere ottenuta solo attraverso la fovea, la piccola zona centrale della retina che ha la piú elevata concentrazione di fotorecettori. Perciò gli occhi sono costretti a muoversi allo scopo di fornire informazioni sugli oggetti che devono essere osservati con una certa precisione (tranne quando si tratti di oggetti molto piccoli rispetto all'angolo da essi sotteso nel campo visivo). I muscoli oculomotori, sotto il controllo del cervello, dirigono gli occhi nel punto significativo (si veda l'articolo di E. Llewellyn Thomas *I movimenti dell'occhio*, in « Le Scienze », n. 11, luglio 1969).

Durante la visione normale di oggetti immobili, l'azione degli occhi si alterna tra fissazioni, quando lo sguardo è fisso in un punto del campo visivo, e rapidi movimenti oculari. Ognuno di questi movimenti porta a fissare lo sguardo su un altro punto del campo visivo. Normalmente si hanno due o tre movimenti al secondo ed essi sono cosí rapidi che occupano soltanto il 10 % circa del tempo di visione.

L'apprendimento visivo e il riconoscimento implicano un immagazzinamento nella memoria e una capacità di recupero. Attraverso il cristallino, la retina e il nervo ottico le cellule nervose situate nei centri ottici della corteccia cerebrale vengono attivate e a questo livello si forma un'immagine dell'oggetto osservato. (L'immagine, naturalmente, è modellata dall'attività neurale ed è del tutto diversa dall'immagine retinica dell'oggetto stesso). Il sistema mnemonico del cervello deve contenere in sé la rappresentazione interna di ogni oggetto di cui sia necessario il riconoscimento. Imparare a conoscere un oggetto e acquistare familiarità con esso costituisce il processo di costruzione di questa rappresentazione interna. Il riconoscimento dell'oggetto stesso, quando lo si incontra per la seconda volta, è il risultato della sua messa a confronto con la rappresentazione interna di esso già contenuta nella memoria.

Non sappiamo con sicurezza se il riconoscimento visivo costituisca un processo parallelo e immediato, o viceversa un'operazione sequenziale di montaggio. Gli psicologi della scuola della *Gestalt* hanno sostenuto che gli oggetti sono riconosciuti come un tutto globale senza che sia necessaria una analisi delle loro parti componenti. Secondo questa tesi la rappresentazione interna di ciascun oggetto sarebbe un tutto unitario che viene messo a confronto con l'oggetto in una singola operazione. Piú recentemente altri psicologi hanno avanzato l'ipotesi che la rappresentazione interna sia come un mosaico, ovvero una composizione di diversi pezzi o elementi. Durante il riconoscimento questi elementi sono confrontati consecutivamente con quelli corrispondenti dell'oggetto da riconoscere. Il riconoscimento è completo quando tutti i pezzi del mosaico sono stati confrontati con successo.

L'ipotesi del riconoscimento seriale è convalidata principalmente dai risultati di esperienze che misurano il tempo impiegato da un soggetto per riconoscere oggetti diversi. L'esperimento tipico consiste nel fare esaminare a un soggetto una serie di oggetti (in genere figure astratte) con il compito di individuare un oggetto « campione » precedentemente immagazzinato nella memoria. Il tempo che il soggetto impiega a considerare ciascun oggetto (o per riconoscerlo come quello campione o per scartarlo) viene misurato. Questo tempo è in genere molto breve, ma è possibile misurarlo in vari modi con notevole accuratezza. Tutti gli oggetti sono abbastanza piccoli da poter essere riconosciuti con un'unica fissazione, cosí che i movimenti oculari non entrano in gioco e non influenzano il tempo necessario per il riconoscimento.

Gli esperimenti di questo tipo hanno portato a due risultati di carattere generale. In primo luogo, si è osservato che in media un soggetto impiega piú tempo a riconoscere un oggetto che a scartarlo. Questo è quanto ci si potrebbe aspettare qualora gli oggetti venissero riconosciuti gradualmente, elemento per elemento. Quando un oggetto viene confrontato mentalmente con la rappresentazione interna dell'oggetto

campione, se esso è diverso per qualche particolare risulterà non corrispondente alla rappresentazione interna e verrà pertanto scartato senza procedere all'ulteriore controllo di tutti gli altri elementi, mentre se l'oggetto è uguale al campione richiederà il confronto di tutte le sue parti componenti. I risultati sembrano contraddire l'ipotesi degli psicologi della *Gestalt* di una rappresentazione interna globale messa a confronto con l'oggetto con un'unica operazione. È presumibile che, in questo caso, il soggetto impiegherebbe lo stesso tempo a riconoscere e a scartare un oggetto.

Il secondo risultato è stato ottenuto variando il grado di complessità dell'oggetto campione immagazzinato nella memoria. È stato osservato che i soggetti impiegano piú tempo a riconoscere oggetti complessi che a riconoscere quelli piú semplici. Anche questo risultato è a favore dell'ipotesi del riconoscimento seriale, poiché in un oggetto piú complesso il numero degli elementi da controllare è maggiore. E, per analogia, il risultato sembra anche contraddire l'ipotesi della *Gestalt*.

Non sarebbe giusto dare l'impressione che la natura seriale del riconoscimento di un oggetto sia un dato di fatto definitamente accertato a spese del concetto unitario sostenuto dagli psicologi della *Gestalt*. Questi hanno dimostrato in modo convincente che un oggetto può avere realmente una qualche forma di « unità primaria », cosicché è spesso possibile individuarlo come entità separata anche prima che abbia inizio un vero e proprio processo di riconoscimento. Per giunta alcuni degli esperimenti sopra descritti per misurare il tempo di riconoscimento dimostrano che, per lo meno nel caso di oggetti molto semplici, quando un oggetto diviene molto familiare la sua rappresentazione interna diventa piú globale e il suo processo di riconoscimento corrispondentemente piú parallelo. Ciò nondimeno il peso delle prove sembra appoggiare l'ipotesi seriale, per lo meno quando si tratta di oggetti non molto semplici e familiari.

Se la rappresentazione interna di un oggetto nella memoria è un mosaico di pezzi, due interrogativi sorgono spontaneamente: quali sono questi «pezzi », cioè quali sono le parti di un oggetto selezionate dal cervello come elementi chiave per l'identificazione dello stesso? E inoltre, in che modo tali elementi vengono integrati e messi in rapporto fra loro per formare la completa rappresentazione interna dell'oggetto? Lo studio dei movimenti oculari durante la percezione visiva è fonte preziosa di informazioni su questi due punti.

Negli esperimenti relativi al primo punto il procedimento in generale consiste nel presentare al soggetto una figura o un oggetto cosí grande e vicino agli occhi da non poter essere registrato sulle fovee senza movimenti oculari. Per esempio, una figura larga 35 cm, e distante dagli occhi 100 cm, che sottende un angolo orizzontale di 20° per ogni occhio – all'incirca l'angolo sotteso da una pagina di questa rivista tenuta alla distanza di un braccio.

In queste condizioni il soggetto deve muovere gli occhi e osservare in lungo e in largo la figura, fissandone tutti i punti che desidera vedere con chiarezza. L'assunto è che egli guardi soprattutto quelle parti della figura che considera piú significative; cioè i punti che gli forniscono la maggior quantità di informazione sulla figura. Essi vengono localizzati per tentativi con la visione periferica e poi fissati direttamente per un'ispezione dettagliata. (È importante osservare che in questi esperimenti, e negli altri che descriveremo, il soggetto riceve soltanto istruzioni di carattere generale come « guarda la figura », o non ne riceve addirittura nessuna. Se gli vengono impartite istruzioni piú precise, per esempio di osservare e descrivere alcuni specifici aspetti della figura, il soggetto in genere fissa direttamente la parte indicata, come è facilmente prevedibile).

Quando i soggetti guardano liberamente in queste condizioni delle figure semplici, per esempio disegni lineari, si è rilevato che i loro punti di fissazione tendono a raggrupparsi a grappolo intorno agli angoli della figura. Per esempio, Leonard Zusne e Kenneth M. Michels hanno effettuato un esperimento di questo tipo alla Purdue University, adoperando come figure dei disegni lineari di poligoni semplici. Dai punti di fissazione dei loro soggetti nell'osservare queste figure risultò evidente che gli angoli attiravano gli occhi in misura molto maggiore del resto.

La conclusione che noi suggeriamo è che, almeno per tali figure lineari, gli angoli sono i tratti salienti utilizzati dal cervello per conservare e riconoscere la figura. Che gli angoli di una figura siano la parte piú significativa ai fini della determinazione di una forma è cosa certa. Nel 1954 Fred Attneave III dell'Università dell'Oregon dimostrò come in un disegno lineare siano gli angoli e i punti di massima curvatura le parti piú ricche di informazione. Per illustrare la sua tesi egli presentò una figura ottenuta selezionando i 38 punti di massima curvatura di un'immagine di un gatto addormentato e unendo tali punti con linee rette (*si veda l'illustrazione in alto nella pagina a fronte*). Il risultato è chiaramente riconoscibile.

Una ulteriore prova del fatto che gli angoli e i punti di massima curvatura sono i tratti piú significativi di una figura è stata fornita dagli elettrofisiologi che hanno esaminato l'attività di singole cellule cerebrali. Per esempio, verso la fine degli anni '50, Jerome Y. Lettvin, H.R. Maturana, W.S. McCulloch e W.H. Pitts dell'Istituto di Tecnologia del Massachusetts hanno scoperto nella retina della rana dei neuroni specifici per la identificazione degli angoli. Piú recentemente David H. Hubel e Torsten N. Wiesel della Harvard Medical School hanno esteso questo risultato a gatti e scimmie (in questi animali le cellule specifiche per la identificazione degli angoli sono situate nei centri ottici della corteccia anziché sulla retina). I dati ottenuti dai centri ottici della corteccia umana da Elwin Marg dell'Università della California a Berkeley costituiscono una prima indicazione di come questi risultati possano essere estesi anche all'uomo.

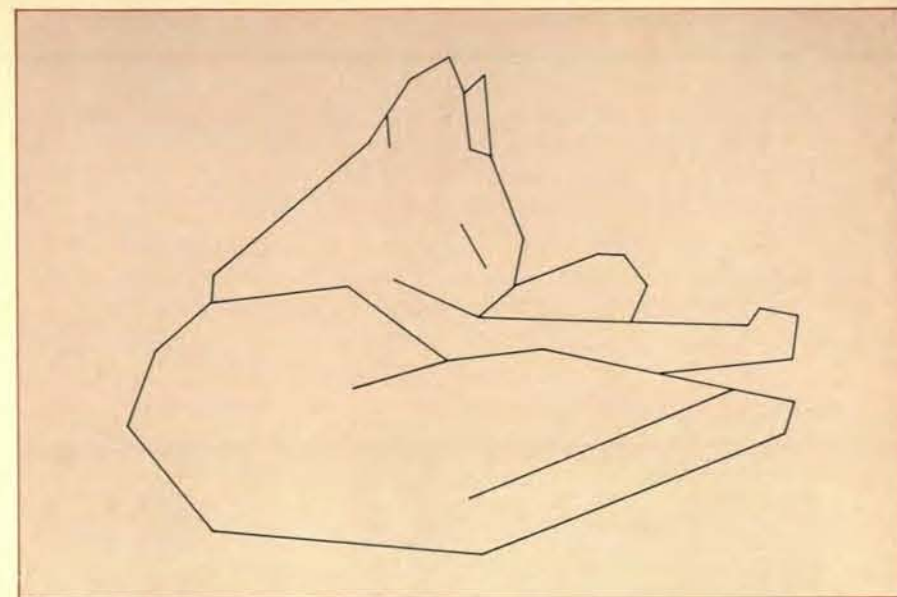
Risultati in un certo senso analoghi sono stati ottenuti con immagini piú complesse di quanto non siano i semplici disegni lineari. Non sorprende che in questi casi anche i tratti caratteristici siano piú complessi. Di conseguenza non si è potuta ottenere una loro descrizione formale. Anche qui, tuttavia, il criterio sembra sia quello di individuare le parti a maggiore contenuto informativo. Norman H. Mackworth e A.J. Morandi hanno effettuato alla Harvard University una serie di registrazioni dei punti di fissazione di soggetti che guardavano due fotografie complesse. Essi conclusero che tali punti si concentravano sui particolari imprevedibili o insoliti, e in particolare sui contorni imprevedibili. Un contorno si può definire imprevedibile quando cambia direzione spesso e irregolarmente e perciò ha un elevato contenuto informativo.

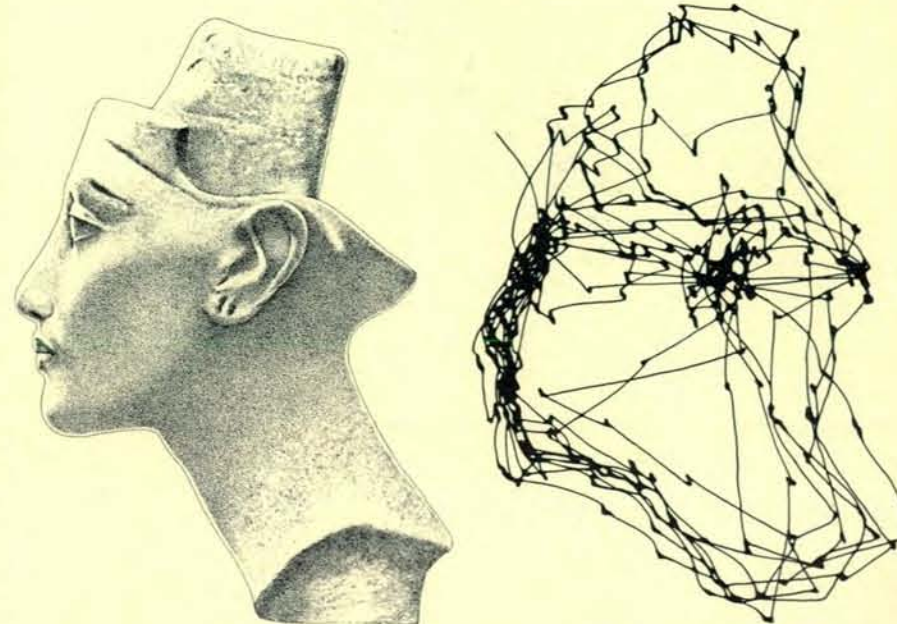
Si può concludere, perciò, che gli angoli e altri particolari ad alto contenuto informativo sono i tratti che il cervello considera piú significativi e che seleziona per ricordare e riconoscere un oggetto.

Il problema successivo riguarda il modo in cui questi elementi vengono integrati dal cervello in un tutto unico – la rappresentazione interna – sicché l'oggetto viene visto come una unità, in altre parole come un oggetto piuttosto che come una sequenza di punti senza rapporto fra loro. Ancora una volta la registrazione dei movimenti oculari ci fornisce molte indicazioni utili. Proprio come lo studio della localizzazione dei punti di fissazione ci ha rivelato la probabile natura dei tratti significativi, cosí l'analisi dell'ordine con cui i singoli punti vengono fissati fornisce uno schema per la integrazione di tali tratti nella rappresentazione interna globale.

L'illustrazione in basso in questa pagina mostra i punti di fissazione di un soggetto durante l'osservazione della fotografia di un busto della regina egiziana Nefertiti. Essa appartiene alla serie di registrazioni effettuate da Alfred L. Yarbus dell'Istituto per i problemi della trasmissione delle informazioni dell'Accademia delle scienze sovietica. La illustrazione mostra chiaramente un aspetto importante dei movimenti oculari durante la percezione visiva, e cioè che l'ordine con cui i singoli punti vengono fissati non è per nulla casuale. I tratti rappresentanti i rapidi movimenti oculari formano delle linee che vanno da un punto all'altro e non in qua e in là per la figura come ci si potrebbe aspettare se gli occhi si posassero ripetutamente sui diversi punti a caso. È chiaro che dopo essersi fissato su un qualsiasi punto, come per esempio l'occhio di Nefertiti, lo sguardo si fissa in genere sull'analogo punto successivo, per esempio la bocca. La registrazione nell'insieme sembra indicare una serie di cicli; in ogni ciclo l'occhio si posa sui tratti caratteristici piú importanti della figura seguendo un percorso abbastanza regolare dall'uno all'altro.

A Berkeley abbiamo recentemente elaborato un'ipotesi sulla percezione visiva che prevede e spiega questa apparente regolarità dei movimenti oculari. In sostanza noi supponiamo che nella rappresentazione interna, o mnemonica, della figura i suoi tratti caratteristici vengano uniti l'uno all'altro in sequenza dalla memoria del movimento oculare richiesto per spostare lo sguardo dall'uno all'altro. Cosí nella scansione della figura gli occhi tenderebbero a muoversi da un punto all'altro seguendo un ordine fisso.

La maggior parte dei dati ottenuti da Yarbus costituisce la somma di numerose fissazioni e non offre una informazione completa sull'ordine con cui esse sono state effettuate. Perciò la regolarità dei movimenti oculari postulata dalla nostra ipotesi non poteva essere confermata con sicurezza in base ai suoi risultati. Per eliminare questa difficoltà e sottoporre la nostra ipotesi a una prova piú specifica, abbiamo compiuto recentemente una nuova serie di registrazioni dei movimenti oculari durante la percezione visiva.

I punti di massima curvatura sono anch'essi importanti come tratti significativi per la identificazione visiva, come è stato dimostrato da Fred Attneave III dell'Università dell'Oregon. L'autore ha selezionato i 38 punti di massima curvatura della figura di un gatto addormentato e li ha uniti con linee rette, eliminando tutte le curve: l'immagine che ne è risultata è ancora facilmente riconoscibile, e questo suggerisce che i punti di curvatura forniscono informazioni utili al cervello nella percezione visiva.

I nostri soggetti guardavano dei disegni lineari di oggetti semplici e di simboli astratti mentre noi misuravamo i loro movimenti oculari (usando delle cellule fotoelettriche per evidenziare i movimenti del « bianco » dell'occhio) registrandoli su nastro magnetico (*si veda l'illustrazione in alto a pagina 26*). Ottenevamo cosí una registrazione permanente dell'ordine con cui i vari punti venivano fissati dai soggetti, e potevamo poi riproiettarla a velocità ridotta analizzandola dettagliatamente per mettere in evidenza la presenza di un andamento ciclico o altre regolarità di movimento. Come negli esperimenti precedenti, i disegni erano abbastanza grandi e vicini agli occhi del soggetto, cioè di nor-

Registrazioni effettuate in un soggetto che guarda la fotografia del busto della regina Nefertiti e che mettono in evidenza delle regolarità nei movimenti oculari. A sinistra è riprodotto il disegno di ciò che il soggetto vedeva; a destra i suoi movimenti oculari registrati da Alfred L. Yarbus dell'Istituto per i problemi della trasmissione della informazione di Mosca. Gli occhi sembrano posarsi sui tratti della testa con una certa periodicità e seguendo un tracciato abbastanza regolare, piuttosto che casuale.

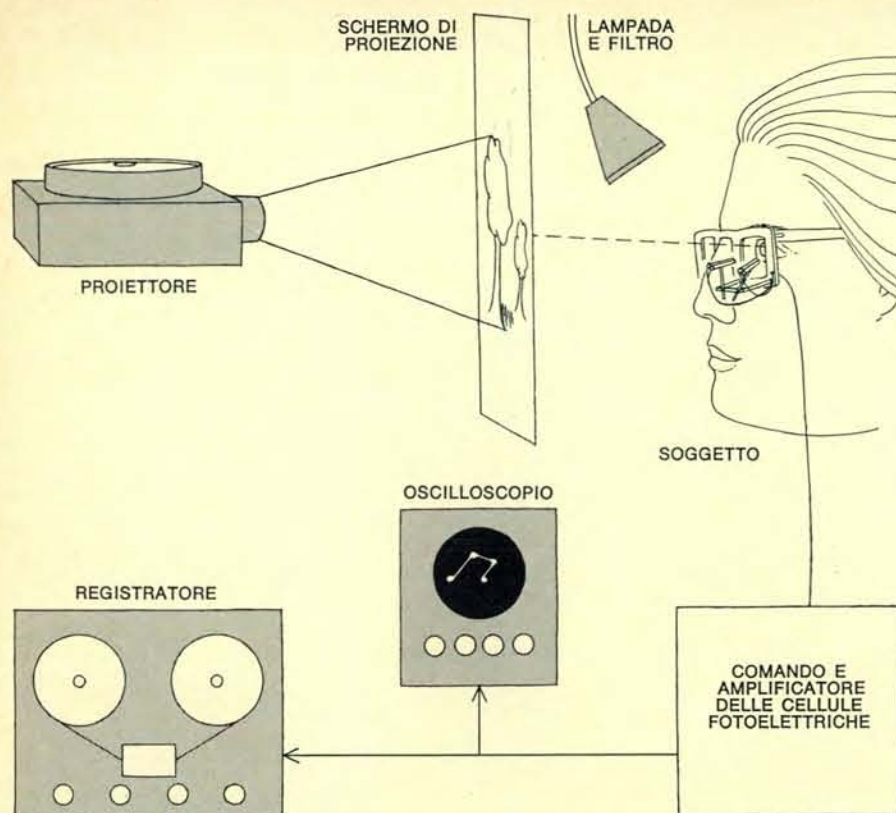



Illustrazione schematica della procedura sperimentale impiegata dagli autori. Il soggetto guarda le figure proiettate su uno schermo in sequenza casuale da un proiettore situato dietro allo schermo. Una luce diffusa infrarossa è posta davanti ai suoi occhi, e i movimenti oculari sono registrati mediante fotocellule, fissate sulla montatura di un paio di occhiali, che rilevano i raggi infrarossi riflessi dal bulbo oculare. I movimenti oculari sono visualizzati su di un oscilloscopio e quindi registrati su nastro.

ma un disegno sottendeva un angolo di circa 20°. Inoltre le figure erano disegnate con linee molto sottili e venivano proiettate a potenza ridotta in modo tale da avere sullo schermo una immagine sbiadita, senza attenuare per nulla la luce normale del laboratorio. In questo modo ottenevamo una immagine di scarsa visibilità e potevamo essere sicuri che il soggetto guardasse direttamente (« fovealmente ») ogni punto che lo interessava, rivelando cosí al nostro apparecchio di registrazione il luogo in cui si fissava la sua attenzione.

I nostri risultati iniziali confermarono largamente la precedente impressione di un andamento ciclico dei movimenti oculari. Abbiamo notato che quando un soggetto guardava una figura in queste condizioni, i suoi occhi, in genere, la esaminavano seguendo – in modo intermittente ma ripetuto – un tracciato fisso che abbiamo definito il suo « tracciato di scansione » per quella particolare figura (*si veda l'illustrazione nella pagina a fronte*). Tra un tracciato di scansione e l'altro erano intercalati dei periodi in cui le fissazioni si susseguivano con minore regolarità.

Ogni tracciato di scansione era caratteristico di un dato soggetto per una data figura. Un soggetto aveva un diverso tracciato di scansione per ogni figura che osservava, e per una data figura ogni soggetto aveva un diverso tracciato di scansione. Un tracciato di scansione tipico per le nostre figure comprendeva circa dieci fissazioni e durava da tre a cinque secondi. I tracciati di scansione occupavano in genere dal 25 al 35 % del tempo di visione del soggetto, mentre il resto era occupato da movimenti oculari meno regolari.

Dobbiamo aggiungere che non sempre era possibile mettere in evidenza un tracciato di scansione. Alcune figure (quella di un telefono, per esempio) sembravano spesso non provocare una risposta ripetitiva, sebbene non fosse possibile individuare delle definite caratteristiche comuni a queste figure. La reazione piú frequente, tuttavia, consisteva nella elaborazione di un tracciato di scansione. A questo punto diventò interessante per noi rifarci alle precedenti registrazioni effettuate da Zusne e Michels, e anche in esse potemmo osservare dei tracciati di scansione che prima erano passati inosservati.

Tale dimostrazione dell'esistenza di tracciati di scansione rafforzava e chiariva le nostre idee sulla percezione visiva. D'accordo con l'ipotesi sequenziale, noi sosteniamo che la rappresentazione interna di un oggetto nella memoria è un mosaico di pezzi. A questo aggiungiamo una ipotesi fondamentale: i «pezzi» sono messi insieme seguendo un caratteristico schema in cui i tratti piú significativi dell'oggetto sono disposti ad anello (*si veda l'illustrazione a pagina 28*). L'anello è costituito da una sequenza di tracce mnemoniche sensoriali e motorie che registrano alternativamente un tratto caratteristico dell'oggetto e il movimento oculare richiesto per raggiungere il tratto successivo. L'anello stabilisce un ordine fisso dei tratti caratteristici e dei movimenti oculari, corrispondente a un tracciato di scansione dell'oggetto.

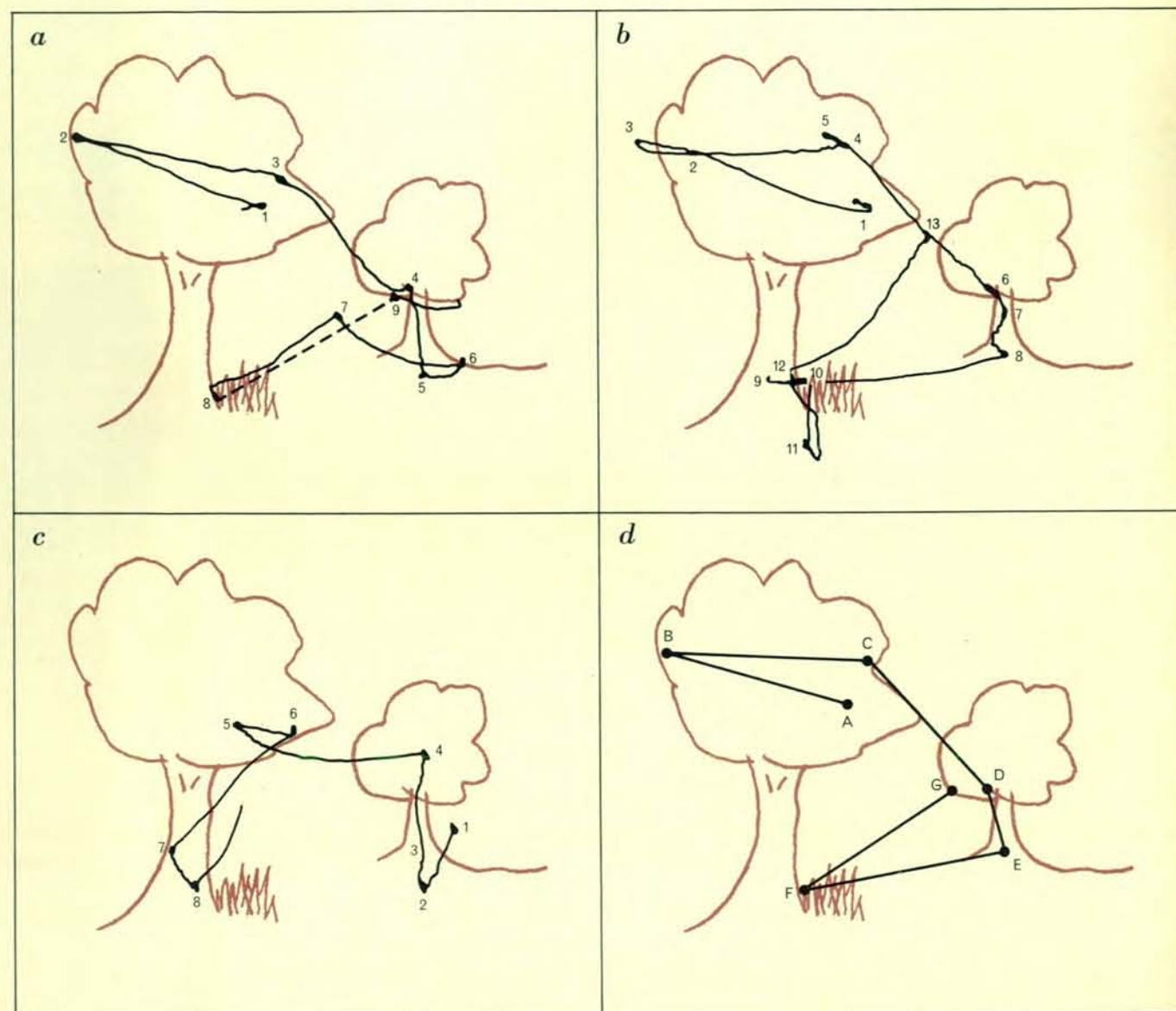
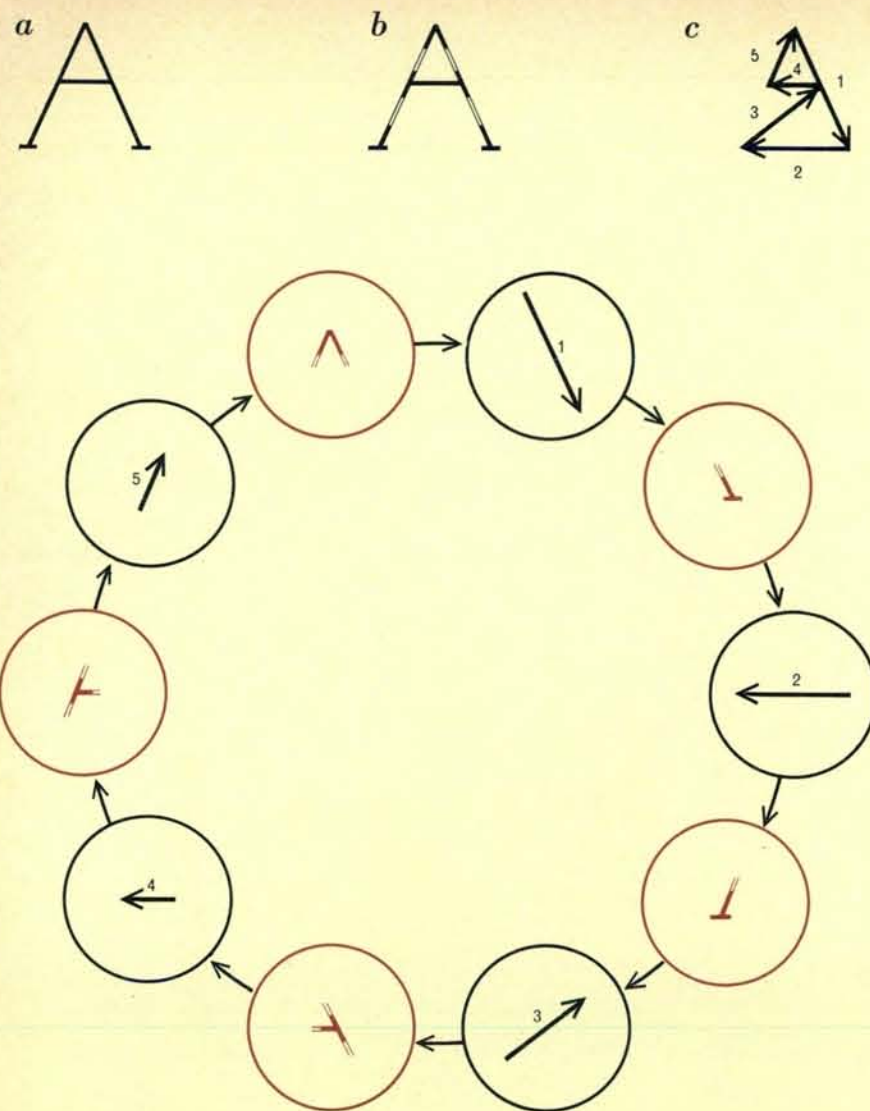
La nostra ipotesi stabilisce che quando un individuo guarda un oggetto per la prima volta e acquista familiarità con esso, lo esamina pezzo per pezzo con i suoi occhi ed elabora un tracciato di scansione per l'oggetto stesso. Nel frattempo egli deposita nella memoria le tracce dello schema anulare che registra tanto l'attività sensoriale quanto quella motoria. Quando, in un secondo tempo, incontra di nuovo il medesimo oggetto, lo riconosce confrontandolo con lo schema anulare che costituisce la rappresentazione interna dell'oggetto stesso nella sua memoria. L'operazione di confronto consiste in una progressiva verifica dei vari tratti caratteristici dell'oggetto e nell'esecuzione dei relativi movimenti oculari, secondo l'ordine dettato dallo schema anulare.

L'ipotesi non soltanto fornisce una spiegazione plausibile sul modo in cui si forma la rappresentazione interna degli oggetti – spiegazione compatibile con l'esistenza dei tracciati di scansione – ma ha anche altre caratteristiche positive. Per esempio, essa ci pone in condizione di stabilire un'interessante analogia tra percezione e comportamento, poiché in entrambi l'attività sensoriale si alterna con quella motoria. Nel caso del comportamento, come quando per esempio si è appresa una sequenza di gesti e la si effettua, la percezione sensoriale di una situazione si alterna con l'attività motoria destinata a realizzare la nuova situazione prevista. Nel caso della percezione (o, piú specificamente, del riconoscimento) di un oggetto, la verifica dei suoi elementi caratteristici si alterna con il movimento degli occhi verso il prossimo elemento previsto.

L'ipotesi dello schema anulare permette anche la previsione e il controllo dei movimenti oculari durante l'operazione di riconoscimento: il succedersi di movimenti oculari e di verifiche dei singoli elementi, essendo dettato dallo schema anulare, dovrebbe ricalcare lo stesso tracciato di scansione che era stato elaborato per quell'oggetto durante la sua osservazione iniziale. La conferma di questa previsione avrebbe ribadito ulteriormente la validità della ipotesi. E poiché era possibile ottenere una conferma sperimentale, decidemmo di fare un esperimento a questo scopo.

L'esperimento comprendeva due fasi, che chiamammo di apprendimento e di riconoscimento. (Con i soggetti, naturalmente non usavamo mai questi termini rivelatori, e dicevamo loro soltanto, come prima, di guardare le figure). Nella fase di apprendimento il soggetto osservava per 20 secondi cinque figure che non aveva mai viste prima. Le figure e le condizioni di osservazione erano simili a quelle del primo esperimento. Per la fase di riconoscimento, che seguiva immediatamente la precedente, le cinque figure erano mescolate con altre cinque completamente sconosciute. Questo al fine di rendere meno facile il compito del riconoscimento. La serie delle dieci figure veniva quindi presentata al soggetto per tre volte in ordine casuale; egli aveva cinque secondi di tempo per guardare ciascuna di esse. I movimenti oculari venivano registrati sia durante la fase di apprendimento sia durante quella di riconoscimento.

Quando analizzammo i tracciati, rilevammo con piacere che le nostre previsioni erano largamente confermate. Durante la fase di apprendimento, nei movimenti oculari del soggetto, compariva un tracciato di scansione e, durante la fase di riconoscimento, i suoi primissimi movimenti oculari nel guardare una figura (presumibilmente durante l'operazione di riconoscimento) ricalcavano in genere lo stesso tracciato che egli aveva elaborato per quella figura nella fase di apprendimento (*si veda la figura a pagina 29*). Secondo la nostra ipotesi, mentre durante la fase di apprendimento si ripetevano i tracciati di scansione e andava formandosi uno schema anulare, nella fase di riconoscimento il soggetto confrontava questo con la figura seguendo il tracciato di scansione dettato dallo schema anulare stesso.

Un altro risultato di questo esperimento fu di mostrare che i diversi soggetti presentano diversi tracciati di scan-

Il tracciato regolare dei movimenti oculari di un soggetto che guarda una data figura è stato definito come « tracciato di scansione » per la figura stessa. In questa figura sono riprodotti due dei cinque tracciati di scansione di un soggetto che guardava il disegno lineare di un albero (*a*, *b*). La linea tratteggiata fra i punti di fissazione 8 e 9 in *a* mostra che la registrazione di questo rapido movimento oculare è stata interrotta da un battito delle palpebre. In *c* si nota come tra un tracciato di scansione e l'altro si manifestino movimenti oculari meno regolari. Il tracciato di scansione del soggetto è mostrato idealizzato in *d*.

Schema anulare proposto dagli autori come modulo per la rappresentazione interna di un oggetto. L'oggetto (*a*) è identificato mediante i suoi tratti piú significativi (*b*) ed è rappresentato nella memoria da questi stessi tratti e dal ricordo del tracciato di scansione (*c*) per mezzo del quale erano stati osservati. L'anello consta quindi di tracce mnemoniche sensoriali (*colore*) che registrano i tratti significativi, e di tracce mnemoniche motorie (*nero*) che rappresentano i movimenti oculari tra un tratto e l'altro.

sione per una stessa figura e, viceversa, un dato soggetto elabora diversi tracciati di scansione per diverse figure (*si veda l'illustrazione a pagina 30*). Questi risultati contribuiscono a scoraggiare altre spiegazioni che potrebbero essere offerte per l'elaborazione del tracciato di scansione. Il fatto che un soggetto abbia tracciati di scansione completamente diversi per diverse figure suggerisce che i tracciati in questione non sono l'espressione di un modo abitudinario di muovere gli occhi quando si osserva ogni figura, come per esempio il leggere il cinese verticalmente, ma piuttosto che essi hanno una origine piú specifica, come uno schema anulare frutto di apprendimento. Analogamente, il fatto che soggetti diversi presentino un diverso tracciato di scansione per la stessa figura ci induce a escludere che i tracciati stessi siano il prodotto di rivelatori periferici degli elementi salienti di una figura che controllerebbero i movimenti oculari indipendentemente dal processo di riconoscimento, in quanto tali rivelatori opererebbero in modo analogo in tutti i soggetti.

Sebbene i risultati del secondo esperimento contribuissero considerevolmente a convalidare le nostre idee sulla percezione visiva, alcuni fatti rimangono oscuri. Per esempio, in certi casi nella fase di apprendimento non viene elaborato nessun tracciato di scansione. In altri il tracciato stesso, evidenziabile nella prima fase, non si presentava nella seconda. In media il tracciato appropriato è comparso nel 65 % circa delle osservazioni effettuate nella fase di riconoscimento. Si tratta di una cifra piuttosto rilevante se si tiene conto dei molti tracciati possibili per ogni figura, ma lascia senza spiegazione l'altro 35 % dei casi.

Probabilmente l'idea fondamentale dello schema anulare ha bisogno di una certa elaborazione. Se ammettiamo la possibilità che tracce mnemoniche registrino altri movimenti oculari intercalati tra i successivi tratti caratteristici della figura, ma non adiacenti nell'anello, e se l'anello originale rappresentasse lo schema di elaborazione preferito e abituale ma non l'unico possibile, la sostituzione occasionale del tracciato di scansione con uno schema abnorme potrebbe trovare una spiegazione (*si veda l'illustrazione in alto a pagina 31*).

Va anche tenuto presente che la registrazione dei movimenti oculari nei nostri esperimenti veniva effettuata mentre i soggetti guardavano figure piuttosto grandi e vicine, il che li costringeva a spostare lo sguardo in qua e in là per la figura per individuarne con chiarezza gli elementi salienti. In una visione piú normale, con una figura o un oggetto abbastanza piccolo da essere tutto visibile con un unico sguardo, non è necessario nessun movimento oculare per il processo di riconoscimento. Noi sosteniamo che in questo caso i successivi stadi della percezione sono paralleli sino al momento in cui un'immagine dell'oggetto si è formata nei centri ottici della corteccia, e che in seguito (come risulta evidente dagli esperimenti effettuati sul tempo di riconoscimento) il confronto fra l'immagine e la sua rappresentazione interna viene effettuato in serie, pezzo per pezzo. Ma in questo caso, dobbiamo postulare anziché dei movimenti oculari da un punto a un altro, una sequenza di spostamenti interni di attenzione con cui si realizza la verifica delle singole parti una dopo l'altra seguendo il tracciato di scansione dettato dallo schema anulare. Cosí a ogni traccia mnemonica motoria nell'anello corrisponde uno spostamento di attenzione che può es-

L'ipotesi dello schema anulare prevede la ricorrenza del tracciato di scansione durante il processo di riconoscimento di un oggetto. Un soggetto guarda una figura tratta da un disegno di Klee (*a*). Mentre si sta familiarizzando con la figura compare un tracciato di scansione (*b*, *c*). Questo si presenta anche (*d*, *e*) durante la fase di riconoscimento ogniqualvolta il soggetto identifica la figura guardando una sequenza di scene conosciute e sconosciute rappresentate in disegni analoghi. Questo particolare tracciato sperimentale di scansione del soggetto è presentato in forma idealizzata in *f*.

sere effettuato o esternamente, sotto forma di un movimento oculare, o internamente, a seconda dell'ampiezza dello spostamento necessario.

Sotto questo aspetto molte delle registrazioni effettuate da Lloyd Kaufman e Whitman Richards all'Istituto di Tecnologia del Massachusetts presentano un notevole interesse. I loro soggetti osservavano figure semplici, come il disegno di un cubo, che potevano essere assimilate con un unico sguardo. In dieci momenti scelti a caso si chiedeva al soggetto di specificare quale punto credeva di guardare. La sua risposta doveva presumibilmente indicare la parte della figura che egli stava cercando di visualizzare. Il suo effettivo punto di fissazione veniva poi registrato in altri dieci momenti scelti a caso (*si veda l'illustrazione inferiore nella pagina a fronte*). I risultati indicano che l'attenzione del soggetto si spostava sulla figura, ma le sue fissazioni rimanevano abbastanza vicine al centro. Questo reperto conferma la teoria che anche gli oggetti di piccole dimensioni vengono esaminati pezzo per pezzo, mediante spostamenti interni di attenzione, pur non essendovi implicati quasi o per nulla i movimenti oculari.

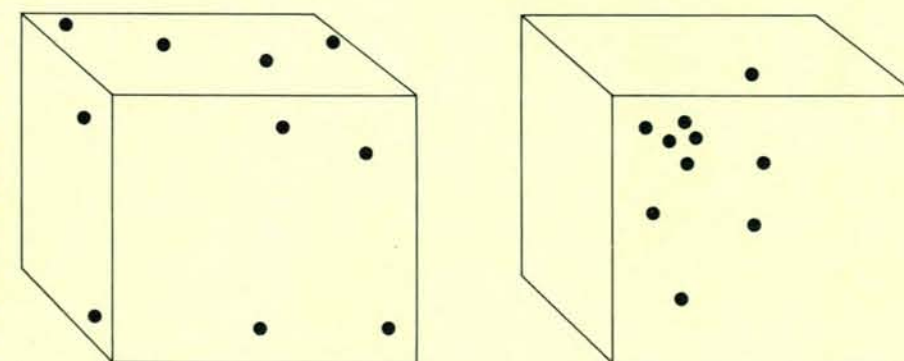
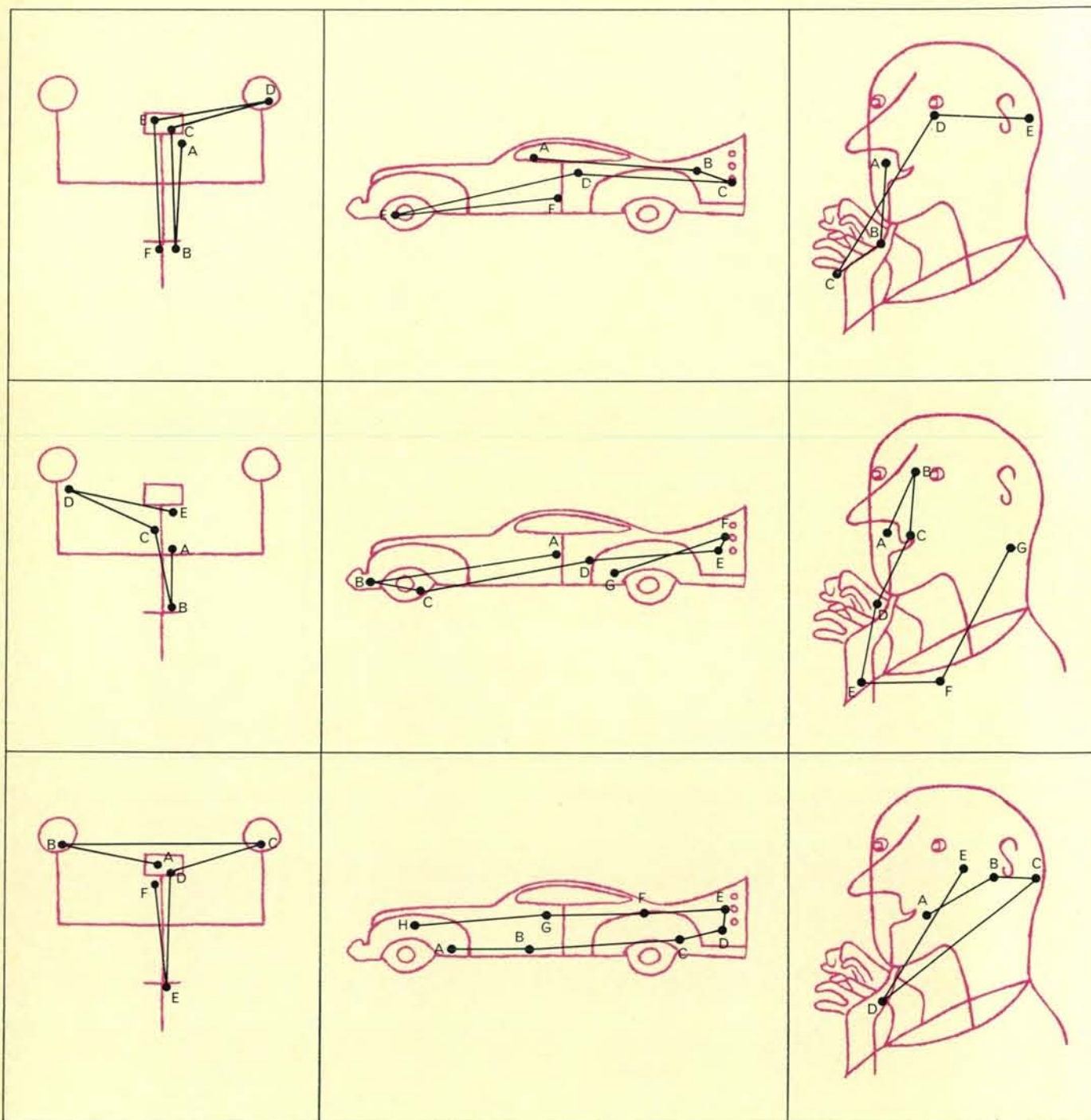
È importante notare, tuttavia, che né questi risultati né i nostri provano che il riconoscimeno degli oggetti e delle figure è necessariamente un procedimento sequenziale in condizioni normali, quando l'oggetto non è cosí grande e cosí vicino da imporre tale gradualità mediante movimenti oculari. Gli esperimenti sul tempo di riconoscimento convalidano questa ipotesi, ma la questione non si può considerare risolta in modo conclusivo. La situazione sperimentale da noi creata imponeva al soggetto di guardare e riconoscere delle figure gradualmente mediante movimenti oculari; in tal modo si rendeva evidente l'ordine con cui i tratti significativi della figura stessa venivano elaborati, e abbiamo supposto che i risultati potessero essere applicati anche al processo di riconoscimento in condizioni piú normali. I nostri risultati propongono una spiegazione piú particolareggiata del procedimento sequenziale – lo schema anulare che dà luogo al tracciato di scansione – ma questa spiegazione rimane condizionata all'ipotesi della natura seriale del processo.

La figura mostra la varietà dei tracciati di scansione per tre soggetti e tre figure. Ogni striscia orizzontale rappresenta i tracciati elaborati da un soggetto per le tre figure. In verticale invece si vedono i tracciati dei tre soggetti per ogni figura.

Riassumendo, a nostro avviso i risultati sperimentali finora conseguiti inducono a trarre tre conclusioni fondamentali sul riconoscimento visivo di oggetti e figure: 1) la rappresentazione interna o memoria di un oggetto è costituita da un mosaico di pezzi collegati con una operazione di montaggio delle parti piú significative o, a rigore, delle tracce mnemoniche di tali parti: durante il processo di riconoscimento la rappresentazione interna dell'oggetto è confrontata sistematicamente con l'oggetto stesso, pezzo per pezzo; 2) i tratti significativi di un oggetto (come gli angoli o le curve nei disegni lineari) sono le parti di esso che hanno il piú elevato contenuto informativo; 3) le tracce mnemoniche che registrano i tratti significativi vengono « montate » sino a dar luogo alla rappresentazione interna globale mediante la connessione con altre tracce mnemoniche che registrano gli spostamenti di attenzione necessari per passare da un pezzo a un altro, sia che questo passaggio avvenga mediante movimenti oculari, sia che si richiedano solo spostamenti interni dell'attenzione; gli spostamenti di attenzione collegano i tratti significativi in un ordine preferenziale, con formazione di uno schema anulare e produzione di un tracciato di scansione, tracciato che viene generalmente ricalcato durante l'operazione di riconoscimento, quando viene effettuato il confronto dei singoli pezzi.

Queste conclusioni esprimono chiaramente una concezione sequenziale dell'apprendimento visivo e del riconoscimento. Nella tendenza a orientamenti di questo genere, in vista di una migliore comprensione della percezione visiva, è manifesta l'influenza della produzione corrente sul riconoscimento computerizzato di forme in cui l'approccio seriale gode da molto tempo i favori dei ricercatori. In realtà, i calcolatori e i concetti relativi all'elaborazione delle informazioni, in genere sequenziali in natura, stanno avendo una influenza sempre piú grande nelle ricerche sul cervello in generale.

Le nostre teorie sul processo di riconoscimento visivo offrono un contributo in questo settore. Le abbiamo elaborate contemporaneamente a un sistema analogo per il riconoscimento computerizzato di forme. Sebbene questo non sia stato utilizzato in pratica, uno schema in certo senso analogo è adoperato nel sistema di riconoscimento visivo di un robot allo studio di un gruppo di ricercatori allo Stanford Research Institute. Noi crediamo che questa fruttuosa collaborazione fra biologia e tecnica debba continuare a lungo a completo vantaggio di entrambe.

Uno schema anulare modificato considera anche movimenti oculari meno regolari che non seguono il tracciato di scansione. Diversi movimenti, che comparivano nel 35 % delle osservazioni in fase di riconoscimento, sono situati nella parte centrale di questo anello. L'anello esterno, che consta di tracce mnemoniche sensoriali (*colore*) e motorie (*nero*), rappresenta il tracciato di scansione e l'ordine preferenziale di elaborazione.

Spostamenti interni di attenzione sostituiscono apparentemente i movimenti oculari nel caso di oggetti abbastanza piccoli da poter essere esaminati con un'unica fissazione. La attenzione del soggetto, secondo le indicazioni da lui fornite, si sposta in qua e in là spaziando per la figura (*a sinistra*), mentre i punti di fissazione registrati (*a destra*) si mantengono abbastanza stazionari. L'illustrazione è stata ricavata da un lavoro di Lloyd Kaufman e Whitman Richards dell'Istituto di Tecnologia del Massachusetts.

Le fibre elastiche nei tessuti

Queste fibre permettono a tessuti come la pelle, le arterie e i legamenti di stirarsi e di rilasciarsi. Due loro componenti sono stati separati, e si è in via di stabilire la loro composizione e le modalità della loro sintesi

di Russell Ross e Paul Bornstein

Con l'evolversi degli animali da forme unicellulari a forme multicellulari è sorta la necessità di collegare tra loro le cellule dei vari tessuti, e il processo di selezione naturale ha portato allo sviluppo di un complesso connettivo intracellulare. Negli animali superiori le caratteristiche di questo tessuto riflettono le proprietà funzionali di ciascun organo. Per esempio, al pulsare del sangue in un'arteria, la parete arteriosa si stende e si rilascia alternativamente in modo simile alla gomma. Le pareti degli alveoli, i sacchi aerei dei polmoni, si comportano allo stesso modo, espandendosi durante l'inspirazione e rilasciandosi nel corso dell'espirazione. Anche tessuti come la pelle e i tendini sono elastici. L'elasticità di questi vari tessuti è dovuta a una fibra biologica che si comporta come la gomma, ma che ha proprietà fisiche e chimiche assai caratteristiche. Questa fibra elastica, uno dei principali componenti del tessuto connettivo già citato, è attualmente oggetto di intensi studi. La conoscenza dettagliata della costituzione e del comportamento delle fibre elastiche presenta un interesse clinico oltre a un interesse puramente scientifico, tanto piú che alcune malattie, e lo stesso processo di invecchiamento, implicano una perdità di elasticità del tessuto connettivo.

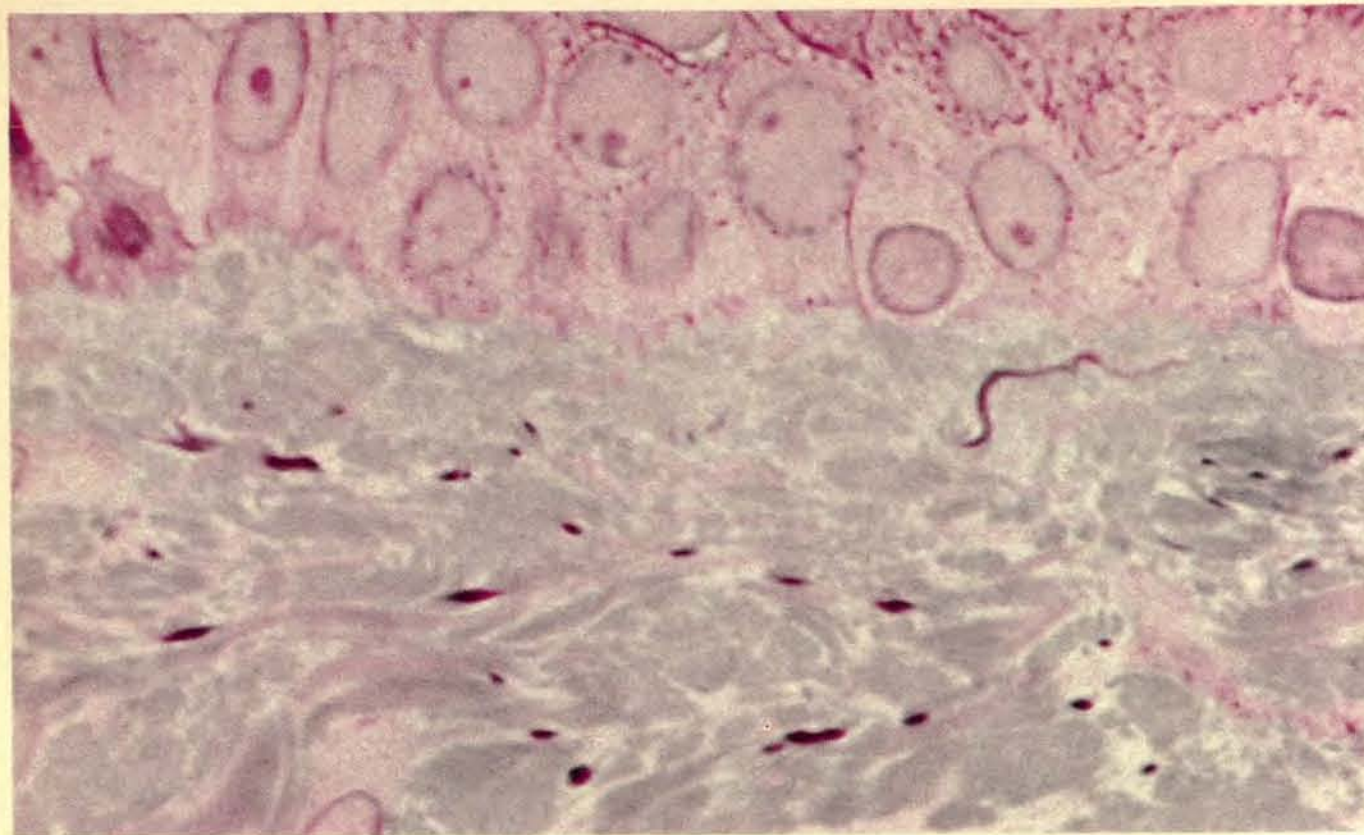
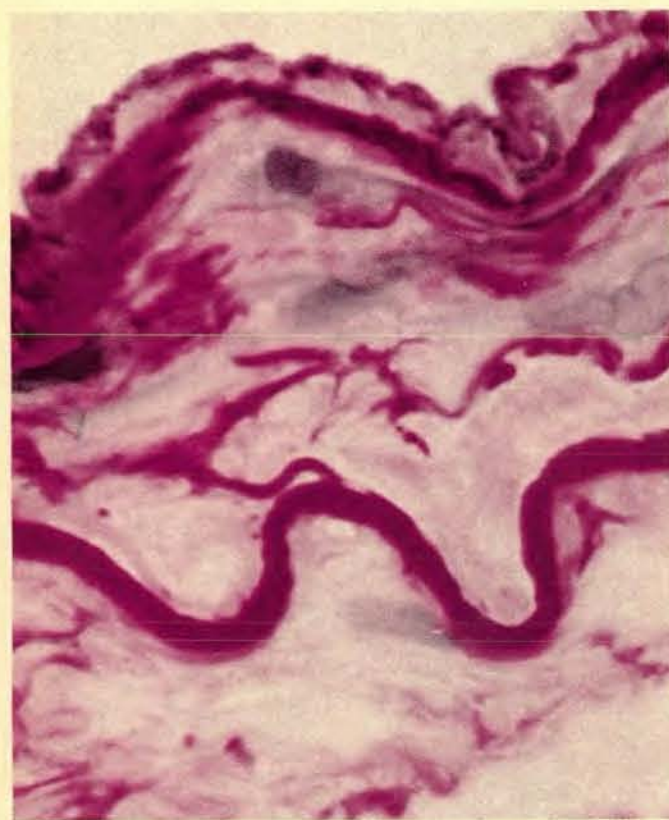
I principali componenti dei tessuti elastici sono il collageno, una proteina fibrosa che presenta una considerevole resistenza alla trazione, e le fibre elastiche che danno al tessuto la capacità di distendersi. Come la gomma, le fibre elastiche si allungano facilmente e rapidamente con una minima perdita di energia in calore, presentano una moderata forza elastica di richiamo quando vengono distese completamente e ritornano rapidamente alle loro dimensioni originali quando cessa la sollecitazione. A rigor di termini, la elasticità è definita come resistenza alla deformazione, e in questo senso, materiali elastici sono sostanze come l'acciaio e il vetro. Nell'uso comune, tuttavia, il termine « elastico » è applicato a materiali che hanno un basso coefficiente di elasticità (cioè possono essere deformati da una piccola forza) e che reversibilmente riprendono le loro dimensioni e forma originali anche in seguito a notevole deformazione.

Esaminando al microscopio ottico un tessuto elastico opportunamente preparato, si possono vedere chiaramente le fibre elastiche. In un tessuto ricco di fibre collagene, come i legamenti, le fibre elastiche sono cilindriche e si intrecciano tra i fasci delle fibre collagene. In una parete arteriosa le fibre elastiche formano strati concentrici interrotti da canali che consentono alle cellule muscolari della parete di comunicare attraverso gli strati.

Quando si poté disporre del microscopio elettronico per uno studio piú dettagliato delle fibre elastiche, si sperò di ottenere delle informazioni sulla struttura delle fibre stesse, il che avrebbe contribuito alla comprensione delle loro proprietà elastiche. I primi studi furono poco incoraggianti. Le fibre apparivano semplicemente come delle masse amorfe di materiale senza alcuna struttura distinta. Tuttavia con il miglioramento della preparazione dei campioni e del potere risolvente del microscopio elettronico, fu possibile accertare che le fibre elastiche hanno una struttura in cui compaiono due componenti. Uno era una microfibrilla con un diametro di circa 110 angstrom; le fotografie ottenute mostravano aggregati di queste microfibrille ammassati insieme nella fibra. L'altro componente era invece amorfo. Quando in seguito fu possibile isolare e analizzare il materiale amorfo, si trovò che era composto per la maggior parte da una proteina assai particolare chiamata elastina. Questa proteina era già stata studiata precedentemente; essa è costituita da una particolare sequenza di amminoacidi, è notevolmente insolubile e ha alcune altre proprietà caratteristiche.

Alla Facoltà di medicina dell'Università di Washington abbiamo studiato la natura di questi due componenti. La nostra ricerca è basata su studi relativi allo sviluppo delle fibre elastiche in animali dallo stadio fetale a quello adulto. Usando il microscopio elettronico per seguire la formazione dei legamenti e dei tendini durante l'embriogenesi, non abbiamo trovato alcuna traccia del componente amorfo nelle fibre elastiche embrionali. A questo stadio le fibre sembravano contenere soltanto le microfibrille. Le proprietà di colorazione delle microfibrille suggerivano che sulla loro superficie fossero distribuite delle cariche negative. Esse mostravano un'altra caratteristica interessante: quando le microfibrille cominciavano a fare la loro comparsa all'esterno della cellula, esse si trovavano molto vicine a questa, spesso occupando delle piccole invaginazioni o cavità della membrana cellulare. Poiché i gruppi di microfibrille durante questo periodo prendevano una forma che apparentemente determinava la forma definitiva delle fibre elastiche adulte, sembrava che lo sviluppo delle microfibrille, sebbene queste fossero all'esterno, fosse in qualche modo sotto il controllo della cellula (*si veda la figura a pagina 35*).

Con il procedere dello sviluppo, il componente amorfo cominciava a com-

Microfotografia delle fibre elastiche contenute nella parete di un'arteria in sviluppo di una giovane cavia, ingrandite circa 80 000 volte. Una fibra elastica è visibile in sezione longitudinale (*da destra in alto, a sinistra in basso*); altre sono visibili in sezione obliqua (*a destra al centro*) e in sezione trasversale (*in basso a destra*). Ciascuna fibra è costituita da fasci di microfibrille che circondano l'elastina e s'infiltrano in mezzo a questa proteina, che ha aspetto amorfo. La parte sinistra della microfotografia è occupata da un altro componente del tessuto elastico, il collageno, sotto forma di fibrille. La maggior parte delle fibrille sono visibili in sezione longitudinale, alcune in sezione trasversale.

La pelle è uno dei tessuti dotati di elasticità per la presenza di fibre elastiche. In questa sezione, ingrandita 500 volte, sono visibili le cellule dell'epidermide al di sotto delle quali si trova il tessuto connettivo, costituito in prevalenza da fibre collagene (*colore chiaro*). Le fibre elastiche nel tessuto connettivo si distinguono perché sono colorate in rosso porpora intenso.

Un'arteria elastica, l'aorta, presenta anelli concentrici di fibre elastiche (*rosso-porpora*) separate da cellule (*rosa con nuclei scuri*) e fibre collagene. L'ingrandimento è di 2000 volte.

L'arteria coronaria, è molto più stretta dell'aorta. Ha un singolo strato di tessuto elastico prossimo alla superficie interna. Le cellule muscolari sono separate da fibre di collageno.

parire nelle fibre elastiche, e al momento in cui l'animale aveva raggiunto l'età adulta queste fibre erano formate in prevalenza da elastina, mentre le microfibrille costituivano soltanto una parte relativamente piccola della fibra sviluppata.

A questo punto della ricerca sorsero molti problemi biochimici. Quale era la relazione tra le microfibrille e il materiale amorfo? La proteina microfibrillare era un precursore dell'elastina o non aveva alcun rapporto con questa? E come veniva sintetizzata l'elastina? È abbastanza ben provato che le proteine formate all'esterno delle cellule sono costituite da unità solubili relativamente piccole che sono state sintetizzate su ribosomi nell'interno della cellula e quindi trasportate all'esterno dove avviene il montaggio finale della macromolecola. In che modo tali unità solubili vengono riunite nell'elastina, altamente insolubile? E quale tipo di legame chimico viene utilizzato per formare questa proteina?

Molti ricercatori hanno preso in considerazione delle ipotesi sulla natura della molecola dell'elastina indotte da analogie di comportamento con la gomma, che ha una struttura molecolare nota. Le prove di allungamento sotto sforzo effettuate su diversi tessuti elastici e fibre elastiche isolate avevano dimostrato che le curve tensione-deformazione (misura della relazione tra la deformazione e la sollecitazione impressa, o carico) sono molto simili alle curve ottenute con la gomma naturale leggermente vulcanizzata. Questo fatto suggeriva che l'elastina potesse avere le stesse caratteristiche strutturali della gomma. Ora, nella gomma vulcanizzata lunghe catene idrocarburiche sono unite insieme in una struttura tridimensionale da ponti trasversali ampiamente spaziati. Tra le catene non c'è alcuna significativa interazione di attrazione o di repulsione. Allo stato naturale (in assenza di sollecitazione) le catene sono ripiegate in modo irregolare. Quando il materiale viene stirato, le catene si allineano, assumendo gradualmente una disposizione cristallina, e questo orientamento delle catene è facilitato dall'assenza di forze tra catena e catena. Tuttavia i legami trasversali limitano l'entità dello scorrimento delle catene e questo rende ragione della forza elastica della gomma. Quando la sollecitazione viene meno le catene ritornano alla loro disposizione irregolare – che è il loro stato naturale di maggiore entropia.

Per analogia, quindi, si potrebbe supporre che l'elastina, come la gomma, sia composta da lunghe catene molecolari legate da scarsi ponti trasversali, e prive di qualsiasi conformazione regolare. Questa supposizione è in accordo con l'aspetto amorfo dell'elastina; perfino con l'esame ai raggi X, come trovò W.T. Astbury dell'Università di Leeds, molti anni fa, l'elastina non mostra alcun segno di una struttura ordinata, né allo stato di riposo, né sotto sforzo. Un'altra osservazione in favore dell'ipotesi che le molecole di elastina costituiscono una struttura relativamente disordinata, è dovuta alla analisi chimica in base alla quale si è stabilito che la maggior parte degli amminoacidi che formano la molecola dell'elastina sono privi di cariche elettriche e idrofobi (idro-repellenti), cosicché, in un ambiente acquoso, qualsiasi interazione tra le molecole deve essere debole.

Alcune recenti ricerche, tuttavia, hanno dimostrato che l'elastina ha una struttura più complessa di quanto non si fosse supposto. Torkel Weis-Fogh dell'Università di Cambridge ha trovato che lo stiramento e il rilassamento dell'elastina è accompagnato da variazioni dello stato energetico della molecola, variazioni che possono essere attribuite a interazioni superficiali tra l'acqua dei tessuti e la parte centrale della molecola dell'elastina, che è fortemente idrofoba e parzialmente strutturata.

Per la sua inerzia chimica e la sua resistenza alla idrolisi con acidi e alcali deboli, l'elastina può essere estratta dal tessuto elastico in forma abbastanza pura come residuo insolubile. Mediante lungo trattamento in autoclave, in acqua, o mediante estrazione con alcali diluiti a caldo, è possibile ottenere un prodotto che possiede una composizione in amminoacidi ragionevolmente costante e che esaminata al microscopio elettronico mostra il caratteristico aspetto dell'elastina. Tuttavia tale preparato contiene spesso tracce di componenti microfibrillari della fibra elastica e altre impurità, essendo chiaramente impossibile assicurare la rimozione di tutte le sostanze contaminanti con tali procedimenti, senza rischiare di indurre delle alterazioni chimiche di parti delicate della molecola. Quindi ab-

Microfotografie delle microfibrille, ingrandite circa 100 000 volte, in due stadi di sviluppo. Nello stadio più precoce della formazione di una fibra elastica in un legamento di un feto di vitello (*in alto*) si vedono fasci di fibrille di 110 angstrom disposti sopra una cellula. Il secondo componente della fibra, l'elastina, compare più tardi. La sezione trasversale di una fibra tratta da un feto di vitello più maturo (*a destra*) mostra le microfibrille che circondano l'elastina appena formata.

Il disegno schematizza i due stadi di sviluppo di una fibra elastica. La fibra immatura *(a sinistra)* è un aggregato di microfibrille raggruppate in un lungo cilindro disposto lungo una invaginazione della membrana. In uno stadio successivo *(a destra)* una nuova proteina, l'elastina *(in colore)*, sintetizzata sui ribosomi *(punti neri)* del reticolo endoplasmico, è stata secreta dalla cellula nel cilindro formato a microfibrille. La cellula quindi svolge un ruolo determinante nella formazione della fibra.

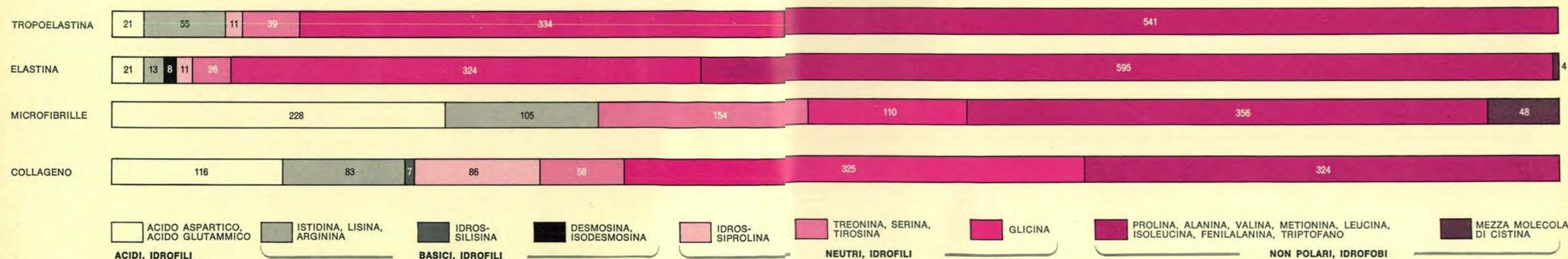
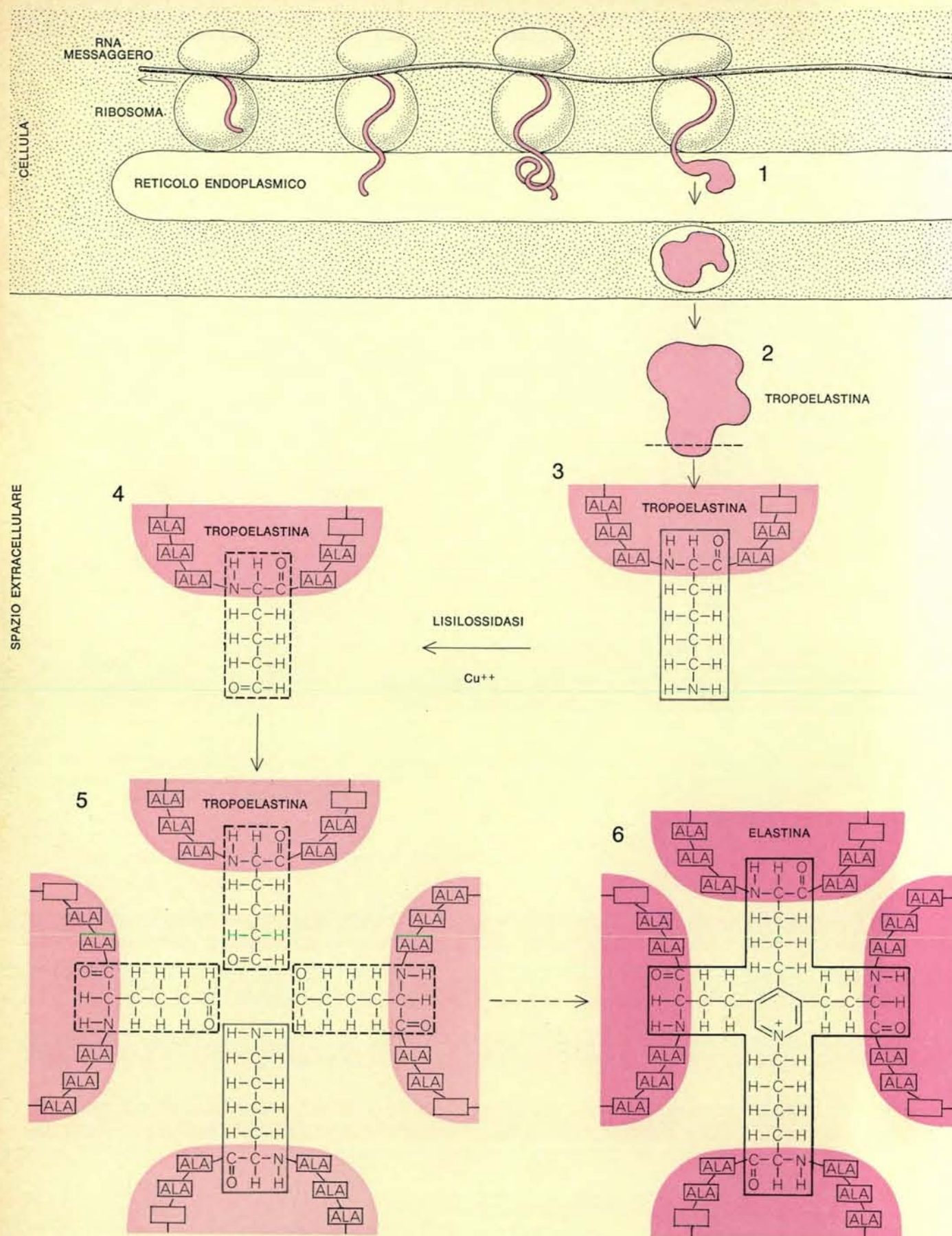
Queste fotografie al microscopio elettronico di preparazioni purificate di fibre elastiche ottenute da un legamento di un feto di vitello mostrano l'efficacia del processo di separazione dei due loro componenti. Nelle fibre intatte *(a sinistra)*, ingrandite 50 000 volte, si trovano le microfibrille e l'elastina dall'aspetto amorfo. Le microfibrille vengono disciolte da un reagente che spezza i ponti disolfurici della proteina, lasciando intatte le fibre elastiche *(a destra)* formate dal solo componente amorfo.

biamo fatto ricorso a un processo in due stadi, piú efficace. Nel primo stadio si isola la fibra elastica, in forma purificata e intatta, mediante l'uso di una sostanza estraente, come una soluzione concentrata di guanidina ed enzimi che demoliscono specificamente il collageno. Poi, per separare le microfibrille dalla elastina, abbiamo utilizzato il fatto che la proteina microfibrillare contiene una grande quantità dell'amminoacido cistina, mentre nell'elastina questo amminoacido è quasi completamente assente. Spezzando il ponte disolfurico (S-S) della cistina, le microfibrille vengono rese solubili e quindi si separano facilmente dall'elastina.

Analizzando le due proteine è risultato che la proteina microfibrillare era notevolmente diversa dall'elastina. Per quanto riguarda la composizione in amminoacidi non c'era alcun rapporto tra l'una e l'altra *(si veda la figura in basso in queste due pagine)* cosicché uno dei nostri problemi aveva una risposta precisa: la proteina delle microfibrille nelle fibre elastiche non è un precursore della elastina.

Veniamo ora a un altro importante problema: come è costruita la molecola dell'elastina? La chiave per la risposta è stata trovata in uno speciale legame trasversale scoperto soltanto recentemente. Fino a una decina di anni fa, l'unico meccanismo noto per stabilire un legame trasversale tra catene proteiche era il ponte disolfurico. Poiché la cistina, l'amminoacido che consente la formazione di questo legame, non è presente in quantità apprezzabile nell'elastina o nel collageno, il problema di quale tipo di ponte trasversale potesse essere implicato nella formazione di entrambe queste sostanze era un enigma.

Il primo indizio fu messo in luce da S.M. Partridge e i suoi collaboratori dell'Agricultural Research Council di Cambridge, Inghilterra. Essi scoprirono che l'elastina incorporava due amminoacidi, prima sconosciuti, che essi denominarono desmosina e isodesmosina (da una radice greca che significa legame). Fu trovato che questi due amminoacidi svolgevano la funzione di stabilire ponti trasversali tra catena e catena nell'elastina. Altri ricercatori dimostrarono che le desmosine venivano sintetizzate a partire da quattro molecole dell'amminoacido lisina e che nell'elastina c'erano altri ponti trasversali derivati dalla lisina che rappresentavano i prodotti di condensazione di due o tre molecole dell'amminoacido.

Varie ricerche sperimentali fornirono dettagli sulle modalità di formazione dei ponti trasversali. Nel 1966 uno di noi (Bornstein), in collaborazione con K.A. Piez dei National Institutes of Health, chiarí le modalità di formazione di tali ponti tra catene in una frazione solubile di collageno. Fu tro-

L'analisi del contenuto degli amminoacidi in elastina, tropoelastina (il suo precursore), microfibrille e collageno permette di stabilire relazioni e differenze tra le quattro proteine. I risultati sono qui riportati come numero di amminoacidi per 1000. (Una certa quantità di asparagina e glutammina, che sono neutre, contribuisce inevitabilmente alla frazione analizzata come acido aspartico e acido glutammico in tutte e quattro le proteine). L'elastina differisce chiaramente dalla proteina microfibrillare: contiene una grossa frazione di amminoacidi idrofobi, una modesta quantità di amminoacidi carichi e poco o nulla cistina. (La cistina che in essa si trova può essere dovuta a contaminazione). La tropoelastina è molto simile alla elastina, come si addice a un precursore. La differenza significativa è che l'elastina contiene desmosina, che è costituita dalla maggior parte della lisina della tropoelastina *(si veda la figura a pagina 38)*; quattro molecole di lisina nella tropoelastina si riuniscono per formare una sola molecola di desmosina nella elastina. Il collageno ha un contenuto in glicina praticamente simile a quello dell'elastina, ma è caratterizzato peraltro da un maggior contenuto di idrossiprolina e, soprattutto, dalla presenza di idrossilisina.

La sintesi dell'elastina implica la formazione di un particolare amminoacido, la desmosina. Il precursore dell'elastina, la tropoelastina, è sintetizzato nella cellula *(1)* ed è quindi secreto nello spazio extracellulare *(2)*. In *3* viene riportato un breve segmento della catena peptidica che costituisce la tropoelastina, contenente una molecola di lisina *(contorno grigio)* circondata da molecole di alanina. Tale lisina è convertita dall'enzima lisilossidasi, in presenza di rame, in un derivato aldeidico *(contorno tratteggiato, 4)*. L'unione di tre molecole di aldeidi con una di lisina *(5)* porta alla desmosina *(contorno nero, 6)*.

vato che la formazione del ponte era dovuta a una reazione tra due aldeidi, prodotte per ossidazione di due molecole di lisina su catene polipeptidiche adiacenti. Fu presto dimostrato da diversi gruppi di ricercatori che anche le fibrille collagene insolubili contenevano ponti trasversali derivati dalla lisina, alcuni dei quali rassomigliavano a quelli identificati nell'elastina. Ulteriori studi mostrarono che la sintesi dei ponti trasversali nella elastina e nel collageno ha inizio a opera di un enzima, la lisilossidasi, che catalizza la formazione delle aldeidi. Si suppone che queste aldeidi fortemente reattive, reagiscono formando un ponte quando vengono a contatto. Nel caso del precursore dell'elastina, che si presume abbia una forma globulare, le quattro unità derivate dalla lisina sembra si uniscano di solito per formare il ponte a quattro braccia della desmosina, che unisce quattro catene di elastina *(si veda la figura nella pagina a fronte)*.

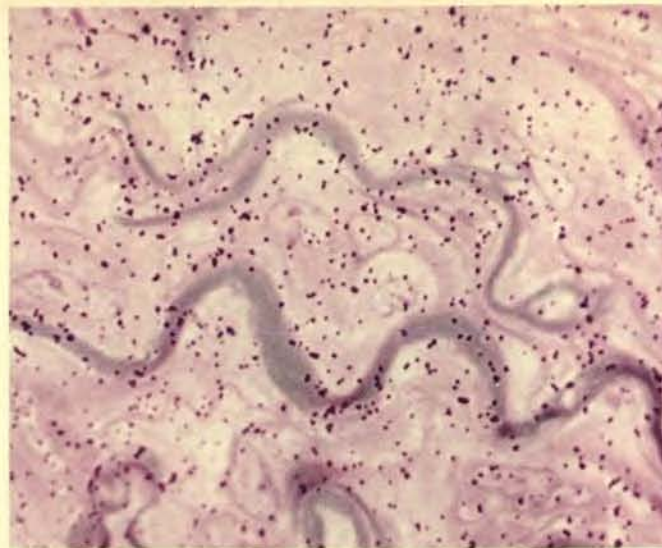
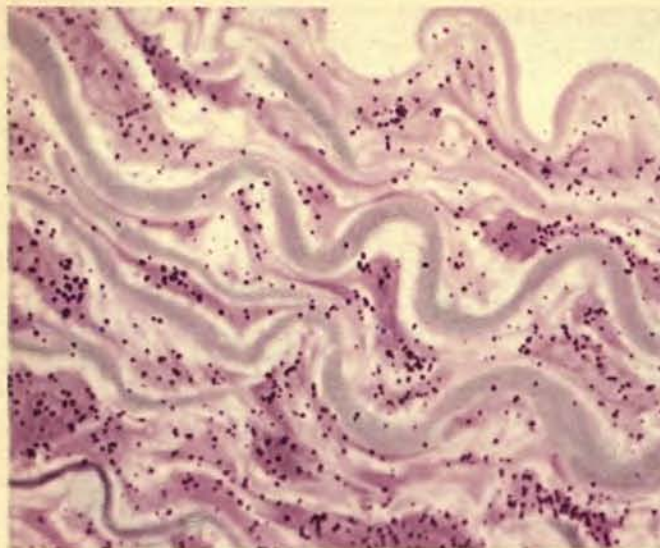
Ulteriori informazioni sulla formazione dell'elastina e sull'identificazione del suo precursore, provengono da studi sui disturbi che derivano dall'inibizione dell'enzima lisilossidasi. Una forma di questa malattia, chiamata osteolatirismo, può essere prodotta sperimentalmente in animali somministrando amminonitrili e composti simili. Nell'osteolatirismo l'inibizione della formazione di ponti trasversali nell'elastina e nel collageno produce delle anomalie strutturali nei tessuti connettivi, particolarmente in quelli dei vasi sanguigni e delle ossa. L'inibizione della formazione dei ponti può essere causata anche da forti deficienze di rame. Negli animali questa deficienza spesso dà luogo alla rottura dell'aorta.

Un gruppo dell'Università dell'Utah ha studiato gli effetti patologici di una marcata deficienza di rame nei maiali giovani. È stato trovato che le fibre elastiche della parete dell'aorta mostravano difetti strutturali poiché contenevano una quantità di elastina notevolmente al di sotto del normale e avevano un accumulo insolitamente grande di microfibrille alla loro periferia. Nel corso di questi studi L.B. Sandberg e i suoi collaboratori estrassero dal tessuto aortico alterato una proteina solubile che mostrava chiaramente le proprietà del precursore dell'elastina a lungo cercato, e che fu quindi chiamata tropoelastina.

La tropoelastina ha una composizione in amminoacidi che è molto simile a quella dell'elastina, con due importanti eccezioni; non contiene desmosina e ha un contenuto molto piú elevato di lisina rispetto all'elastina. Questa situazione è proprio quella che ci si aspetterebbe secondo l'ipotesi generale che il precursore viene convertito in elastina, prodotto terminale insolubile, attraverso la formazione dei ponti di desmosina a partire da lisina. La quantità di lisina trovata nella tropoelastina è in buon accordo con il numero di molecole di lisina richieste per formare i ponti osservati nella elastina. È chiaro che l'aver trovato quantità significative di tropoelastina e una quantità di elastina inferiore al normale nell'aorta di questi maiali con carenza di rame indica che veniva inibita la formazione di ponti fra molecole di protoelastina per la conversione in elastina.

La relazione tra tensione e allungamento di una fibra elastica è stata determinata da Hoffman della Università di Washington e da Mukherjee dei Laboratori di ricerca Goodyear. Un legamento è stato purificato per allontanare i componenti diversi dalle fibre elastiche. La sollecitazione applicata è uguale alla forza applicata parallelamente alla fibra per la sezione trasversale di questa in assenza di sollecitazione. Il rapporto di allungamento è uguale alla lunghezza sotto carico divisa per la lunghezza in condizioni di riposo. Al di sopra di un rapporto 1,03, in pratica, tutte le fibre erano sollecitate. Il comportamento dei campioni era simile a quello della gomma vulcanizzata.

A questo punto possiamo formulare un quadro generale del processo che porta alla produzione di fibre elastiche. La tropoelastina, precursore della elastina, è sintetizzata a livello dei ribosomi, entro la cellula. È composta in modo predominante da amminoacidi (come valina, alanina, glicina e prolina) idrofobi oppure privi di carica, cosí che la molecola tende a essere insolubile in acqua. È possibile, tuttavia, che la tropoelastina contenga abbastanza lisina da formare una sorta di involucro idrofilo (con affinità per l'acqua) in modo che gli amminoacidi idrofobi vengano a trovarsi all'interno della molecola globulare. Questo involucro carico elettricamente e idrofilo può rendere la molecola solubile. Dopo la sua sintesi, la tropoelastina è trasportata fuori dalla cellula e giunge alla superficie di una fibra elastica in via di crescita. Qui, con l'ausilio dell'enzima lisilossidasi, la maggior parte delle lisine nella molecola di tropoelastina vengono convertite in aldeidi, che condensano formando dei ponti tra le varie molecole. Come conseguenza della conversione dei residui lisinici idrofili, la molecola di elastina risultante assume una costituzione prevalentemente idrofoba diventando insolubile. La proteina quindi costituisce un aggregato di subunità che possiede le proprietà chimiche e fisiche necessarie per essere elastica. Non ha la struttura caratteristica della gomma, consistente in catene intrecciate in un modo del tutto irregolare. La sua struttura è invece un aggregato di subunità facilmente deformabili. Queste subunità legate tra di loro funzionano come una massa elastica proprio come un insieme di piccole molle connesse le une con le altre.

Che parte hanno in questo schema le microfibrille della fibra elastica? Il loro ruolo è apparentemente quello di modellare l'elastina in una configurazione fibrosa. Abbiamo già notato che le microfibrille sono le prime strutture che compaiono durante la formazione

Le cellule muscolari liscie sono state identificate come il luogo di sintesi delle fibre elastiche delle arterie mediante autoradiografia. Si somministrava a giovani ratti un amminoacido marcato (prolina) e si preparavano delle sezioni dell'aorta in via di sviluppo dopo 30 minuti (*a sinistra*) e quattro ore (*a destra*). Le sezioni venivano lavate per allontanare la prolina libera ed erano messe a contatto con l'emulsione fotografica. I punti scuri rappresentano i cristalli di bromuro d'argento impressionati dalle radiazioni della prolina incorporata nella proteina. La fotografia a sinistra mostra che dopo 30 minuti la prôteina è presente nelle cellule del muscolo liscio. La fotografia a destra mostra che, dopo quattro ore, la maggior parte della proteina marcata ha lasciato le cellule e si è trasferita nelle fibre elastiche intensamente colorate oppure nelle fibre collagene.

delle fibre elastiche e che formano dei fasci allineati longitudinalmente in cavità vicino alla superficie cellulare. L'elastina non avrebbe di per sé la tendenza a formare una fibra. Si può ritenere che le microfibrille, che sono cariche negativamente, interagiscano con le molecole di tropoelastina, cariche positivamente, e le orientino a formare una fibra durante la formazione dei ponti.

Dobbiamo infine rispondere alla domanda che riguarda il luogo di origine delle fibre elastiche e cioè quali sono le cellule che ne sintetizzano i componenti. Per quanto riguarda le fibre elastiche dei legamenti e dei tendini la risposta è ovvia. La loro formazione deve aver origine nei fibroblasti, le cellule che producono questi tessuti connettivi. Per quanto riguarda le fibre elastiche delle pareti arteriose la risposta non è cosí chiara. È parso ragionevole supporre che il posto piú probabile per la formazione di queste fibre fosse lo strato mediano della parete arteriosa, che è il piú ricco di fibre elastiche e contiene soltanto cellule della muscolatura liscia. Per vedere se queste cellule producevano o meno effettivamente i componenti delle fibre elastiche uno di noi (Ross) e Seymour J. Klebanoff usarono la tecnica dell'autoradiografia, che rende possibile seguire la biosintesi di una grossa molecola seguendo il cammino di un suo componente marcato con isotopi radioattivi. Noi abbiamo usato come tracciante l'amminoacido prolina che è stato somministrato a un ratto in crescita. Seguendo l'incorporazione nelle proteine di questo materiale marcato a opera delle cellule del muscolo liscio, abbiamo trovato che queste cellule sintetizzano effettivamente i tre costituenti maggiori del tessuto connettivo: il collageno, e i due componenti della fibra elastica, le microfibrille e la elastina.

Le anomalie nelle cellule muscolari lisce della parete arteriosa e nei componenti del tessuto connettivo che li circondano hanno un ruolo di primo piano nelle malattie delle arterie, come l'arteriosclerosi. Ulteriori studi su ciascuno di questi componenti dovrebbero essere utili ai fini di comprendere in maniera chiara e definitiva in che modo simili alterazioni nei vasi sanguigni possono portare a malattie vascolari.

In questa autoradiografia eseguita al microscopio elettronico sono localizzati i siti di sintesi nelle cellule del muscolo liscio. La proteina marcata, rappresentata dalle linee ritorte nere, si trova nel reticolo endoplasmatico e nell'apparato di Golgi.

Un'antica città iraniana

5500 anni fa, Tepe Yahyā era un fiorente centro a circa metà strada tra la Mesopotamia e l'India. Avamposto dell'urbanesimo mesopotamico, ebbe una posizione chiave nella diffusione della civiltà da occidente a oriente

di C.C. e Martha Lamberg-Karlovsky

Il regno di Elam e il suo forse più noto vicino, Sumer, furono i due più antichi stati urbani sorti nell'area mesopotamica durante il quarto millennio a.C. Recenti ritrovamenti archeologici dimostrano che il regno elamita comprendeva anche un territorio che si estendeva 800 chilometri più a est. Per più di dieci secoli, a partire dal 3400 a.C. circa, la regione collinosa dell'Iran sudorientale, a circa 100 chilometri dal Mare Arabico, fu il luogo dove fiorì il secondo centro della cultura urbana elamita.

Oggi tutto quanto rimane della città che sorse a metà strada tra l'Eufrate e l'Indo è un grande tumulo di terra che si trova a circa 1400 metri sul livello del mare nella valle di Soghun, 240 chilometri a sud della città di Kerman nella provincia dallo stesso nome. Noto nella regione come Tepe Yahyā, il tumulo è alto 18 metri e ha un diametro di 180. La prima traccia di un insediamento è costituita da un villaggio neolitico che risale a 6500 anni fa, l'ultima è rappresentata da una cittadella della dinastia sassanide che dominò la Persia all'inizio dell'era cristiana. Gli strati intermedi del tumulo provano l'esistenza di legami tra questa orientale città elamita e i centri tradizionali d'occidente.

Questa lunga sequenza archeologica riveste una grande importanza per lo studio dello sviluppo culturale dell'uomo nel suo passaggio da agricoltore a cittadino, ma tre scoperte inaspettate conferiscono a Tepe Yahyā un valore molto maggiore. In primo luogo, l'analisi dei materiali organici associati condotta con il carbonio -14 ha dimostrato che le tavolette di argilla rinvenute in uno degli strati inferiori del tumulo risalgono al 3560 a.C. (± 110 anni). Le tavolette recano incisi caratteri noti come protoelamiti. Le iscrizioni protoelamite e quelle del primo periodo sumerico sono le più antiche scritture conosciute in Mesopotamia, addirittura le più antiche di cui si abbia notizia. Le tavolette di Tepe Yahyā inoltre sono uniche in quanto le prime nel loro genere che è possibile datare in modo preciso. Il ritrovamento di questi esempi di scrittura – antichi quanto i più antichi conosciuti – in un luogo tanto lontano dalla Mesopotamia fu veramente inaspettato.

Il secondo fatto sorprendente è costituito dalla prova che il commercio elamita con il vicino Sumer di materiale insolito come la steatite, una pietra facilmente lavorabile, rappresentava una parte fondamentale nei traffici di Tepe Yahyā. A differenza di Sumer, circondato dalle uniformi pianure alluvionali della bassa Mesopotamia, Elam era un paese collinoso, ricco di risorse naturali. Il commercio elamita forniva ai Sumeri argento, rame, stagno, piombo, pietre preziose, cavalli e materiali più comuni come legno, ossidiana, alabastro, diorite e steatite. La scoperta che questo traffico di steatite si spingeva a oriente fino a Tepe Yahyā conferisce una nuova dimensione alla nostra conoscenza del commercio del quarto millennio.

In terzo luogo, la scoperta di Tepe Yahyā ha contribuito sostanzialmente ad ampliare i confini dell'antico Elam, di cui si aveva una nozione ancora molto confusa. Susa, la più famosa città elamita non è lontana da importanti centri sumeri come Ur e Eridu. Quanto ad altre città elamite citate in alcune iscrizioni (per esempio Awan o Madaktu), la loro ubicazione resta ancora un mistero. La scoperta di una fiorente città elamita come Tepe Yahyā, tanto a oriente della Mesopotamia, costituisce al tempo stesso una sorpresa e una rivelazione. Ciò suggerisce come la civiltà urbana, sorta nella bassa Mesopotamia si aprisse la via verso oriente fino alla valle dell'Indo (nella regione che è ora il Pakistan occidentale).

Sir Aurel Stein, l'esploratore-archeologo britannico, fu il primo a scoprire che l'Iran sudorientale è una regione ricca di importanti resti preistorici. Due siti che Stein sondò brevemente negli anni '30 – Tal-i-Iblis, vicino a Kerman, e Bampur nel Belucistan persiano – sono stati recentemente scavati, il primo da Joseph R. Caldwell dell'Università di Georgia e il secondo da Beatrice de Cardi del Council for British Archaeology. Sebbene si tratti del tumulo più vasto dell'Iran sudorientale, Tepe Yahyā rimase sconosciuto fino all'estate del 1967, quando il nostro gruppo di ricerche del Peabody Museum dell'Università di Harvard lo scoprì durante una spedizione archeologica nella regione.

Attualmente a Tepe Yahyā abbiamo già portato a termine tre stagioni di scavi in collaborazione con l'Iran Archaeological Service, arrivando a determinare una sequenza di sei periodi principali di occupazione. Il sito fu abitato in modo pressoché continuo a partire dalla metà del quinto millennio a.C. fino al 400 d.C. circa. Al termine del periodo elamita, intorno al 2200 a.C., c'è un vuoto di mille anni che non è ancora stato possibile spiegare, ma che trova rispondenza in altri importanti siti dell'Iran: Tepe Yahyā rimase inabitata fino al 1000 a.C., quando vi si insediò una popolazione dell'Età del Ferro.

Il lavoro vero e proprio a Tepe Yahyā cominciò nell'estate del 1968 con una serie di scavi, ciascuno costituito da un quadrato di 9 metri di lato, dalla sommità alla base del tumulo (*si veda la figura nella pagina a fronte*). Vennero quindi fatte piccole trincee di assaggio all'interno della serie di quadrati. Durante la seconda e la terza stagione, gli scavi si estesero orizzontalmente sulla sommità del tumulo e a ovest nella zona esplorata più a fondo. Aprimmo inoltre una trincea a gradini larga 3,6 metri sul lato opposto del tumulo per controllare le sequenze già portate alla luce.

A Tepe Yahyā i più antichi resti di occupazione umana, che in alcuni punti poggiano su terreno vergine, consistono in cinque livelli sovrapposti di costruzioni in mattoni di fango. Li abbiamo attribuiti a un unico periodo culturale – il periodo VI – che l'analisi condotta con il carbonio -14 fa risalire alla metà del quinto millennio a.C. Le strutture del periodo VI sono costituite da una serie di aree, di 1,5 metri di lato, che sembrerebbero luoghi adibiti a magazzino. Per la maggior parte sono prive di porte, e probabilmente vi si accedeva mediante un'apertura praticata nel tetto. I muri sono costituiti da mattoni di fango essiccati al sole, fatti a mano, o da fango impastato (*si veda la figura a pag. 53 in alto*). Frammenti di legno e di stuoie di canne rinvenuti sul pavimento delle stanze sono quanto rimane dei tetti crollati.

Gli utensili del periodo VI comprendono strumenti di osso e di selce. Molte delle selci sono minuscole; tra esse vi sono piccole lame che venivano inserite in un manico d'osso per formare un falcetto. Il vasellame più comune è grossolano e fatto a mano con argilla mescolata a paglia triturata. I vasi, a forma di ciotola e di grandi giare per provviste, presentano decorazioni rosse a linee ondulate o sono interamente dipinte con questo colore. Verso la fine del periodo VI compaiono alcuni pezzi di vasellame più raffinato, di colore giallo, a superficie levigata ottenuta applicandovi uno strato di creta semifluida, oppure di colore rosso con decorazioni nere.

Sotto il pavimento di alcune strutture furono rinvenute sepolture umane, tutte di bambini. Le membra erano state avvicinate al tronco prima dell'inumazione e accanto ai corpi c'erano, intatte, ciotole di fattura grossolana. In una stanza trovammo una statuetta umana il cui viso poggiava su alcuni utensili di selce e d'osso. La scultura di steatite verde scuro, alta 28 centimetri (*si veda la figura nella pagina seguente*), rappresenta chiaramente una figura femminile, tuttavia la sua forma allungata e la presenza di un foro alla sommità del capo suggeriscono una duplice simbologia che raccoglie in sé caratteristiche maschili e femminili.

La cultura neolitica del periodo VI praticava evidentemente l'agricoltura e la pastorizia: tra le ossa di animali identificate ve ne sono di gazzelle selvatiche, bovini, pecore e capre. La presenza di ossa di cammello non basta però a chiarire se in questo periodo così remoto l'animale fosse già stato addomesticato. Tra le piante coltivate vi erano anche i cereali; oggi nell'area di Tepe Yahyā è necessario ricorrere all'irrigazione, ma non si sa se ciò avvenisse anche in tempi neolitici. In ogni caso l'occupazione neolitica del tumulo continuò fino al 3800 a.C. circa.

La transizione dal periodo VI alla Prima Età del Bronzo immediatamente successiva si verificò senza soluzione di continuità. Le strutture del periodo V contengono vasellame grossolano del tipo precedente. La ceramica dipinta, più raffinata, diventa comune e comprende anche nuove tipologie. Uno di questi vasi, la cui superficie era trattata con uno strato di creta semifluida rossa, presenta un motivo decorativo geometrico a zig-zag dipinto in nero. A questo tipo particolare di ceramica nero su rosso abbiamo dato il nome di ceramica Yahyā e alla cultura del periodo V quello di cultura Yahyā.

Gli esempi più comuni di ceramica Yahyā sono costituiti da vasi campaniformi. Spesso questi presentano alla base il marchio del vasaio e finora sono stati individuati sei marchi diversi. La scoperta a Tal-i-Iblis, un sito più vicino a Kerman di circa 160 chilometri, di ceramica dipinta quasi identica con marchi di vasai dello stesso genere, dimostra che il contatto esterno e il commercio erano parte essenziale nella vita

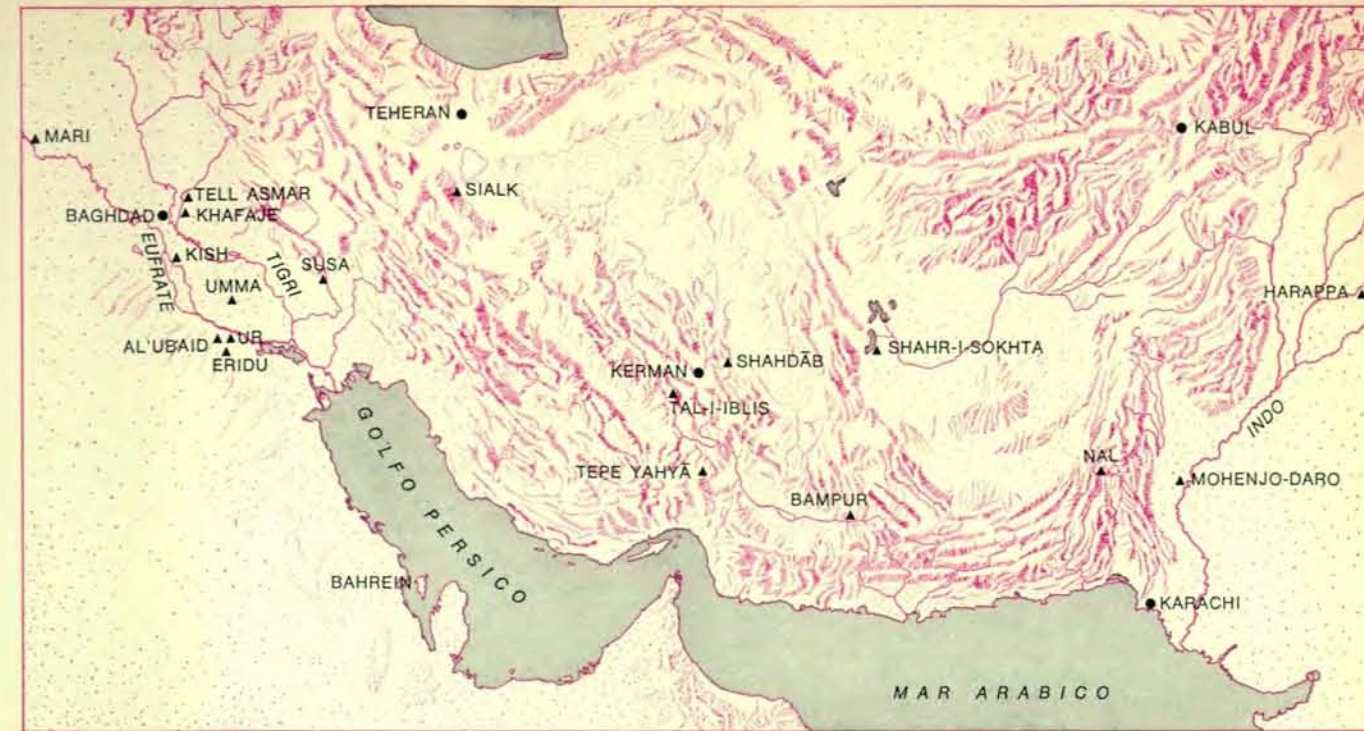
Il grande tumulo di terra, di oltre mezzo chilometro di circonferenza, situato nell'Iran sudorientale e noto nella regione come Tepe Yahyā; il sito venne occupato per la prima volta da una comunità neolitica verso la metà del quinto millennio a.C.

In uno dei magazzini della struttura piú antica di Tepe Yahyā è stata rinvenuta questa statuetta neolitica insieme a utensili di selce e di osso. La scultura rappresenta una doppia immagine: una figura femminile sovrapposta a una stilizzata forma fallica.

della Prima Età del Bronzo a Tepe Yahyā. Tuttavia esistono altre prove di contatti tra regione e regione. La ceramica Yahyā rivela una generica somiglianza con il vasellame dipinto rinvenuto altrove nell'Iran sudorientale, e un tipo di ceramica Yahyā a decorazioni nere su fondo giallo è molto simile a quella di località molto piú a ovest come Bakun. Inoltre i livelli del periodo V di Tepe Yahyā abbondano di materiali importanti. Vi sono utensili di ossidiana, perle di avorio, cormalina e turchese, e vari oggetti di alabastro, marmo e madreperla. Una statuetta di alabastro alta circa 25 centimetri, particolarmente bella, raffigura in modo stilizzato un ariete (*si veda la figura a pag. 56 in alto*). Di nessuno di questi materiali si conoscono fonti locali.

Sebbene l'architettura del periodo V dimostri una continuità con il precedente periodo neolitico, le singole strutture sono piú ampie. Parecchie di esse coprono un'area di 2,5 × 3,5 metri e hanno caratteristiche chiaramente residenziali. Alcune stanze contengono un focolare e un camino. Nei primi livelli i muri sono ancora costruiti con mattoni di fango fabbricati a mano. I mattoni a stampo compaiono verso la metà del periodo V che l'analisi condotta con il carbonio -14 colloca intorno al 3660 a.C. (± 140 anni).

Gli strumenti di bronzo del periodo V, come avvenne ovunque per i primi oggetti di questo metallo, non venivano ottenuti con una lega, ma utilizzando minerali di rame « impuri ». Questa tecnica è tipica dell'antico Sumer dove il minerale grezzo importato da Oman, nella penisola arabica, presentava in natura un alto tenore di nichel. Altrove i primi artigiani del bronzo si servivano di rame grezzo ricco di arsenico. Ceselli, punteruoli, spilli e spatole rinvenuti a Tepe Yahyā sono fatti di questo bronzo a elevata percentuale di arsenico.

R.F. Tylecote e H. McKerrell dell'Università di Newcastle upon Tyne hanno analizzato sei manufatti provenienti da questo sito. Essi scoprirono che il bronzo era stato ottenuto mediante fusione, e ciò dimostra che i fabbri del periodo V riuscivano a produrre le elevate temperature necessarie per trasformare i minerali di rame in metallo fuso. Tuttavia la forma definitiva non veniva ottenuta con una colata, ma mediante una tecnica piú primitiva di forgiatura a caldo e a freddo. Si è trovato che uno degli oggetti, un cesello, contiene il 3,7 % di arsenico, e ciò induce a credere che gli artigiani avessero scelto consapevolmente minerali grezzi ad alto tenore di arsenico. Questa scoperta è un'ulteriore conferma degli scambi commerciali di Tepe Yahyā: per ottenere questo speciale bronzo all'arsenico non si sarebbe potuto ricorrere a nessuno dei giacimenti di rame della regione.

Le prime città sorsero nel regno di Sumer nella bassa Mesopotamia (*a sinistra*). Le piú antiche forme di scrittura conosciute apparvero a Sumer e nel vicino Elam in città come Susa e Sialk. La scoperta di iscrizioni protoelamite a Tepe Yahyā (*al centro*), fa pensare che il commercio tra questa regione e le prime città mesopotamiche abbia provocato il sorgere di città in questa parte dell'antica Persia nel quinto millennio a.C. e abbia condotto al successivo sviluppo della civiltà urbana di Harappa.

Con l'inizio del periodo IV, intorno al 3500 a.C., la comparsa della scrittura a Tepe Yahyā consente di identificare la città come un insediamento protoelamita. Gran parte della ceramica delle prime due fasi di questo periodo, IV-C e IV-B, è quella tipica della precedente cultura Yahyā sia come forma sia come decorazioni. Sebbene ci siano molte prove di contatti esterni, a Tepe Yahyā la transizione al periodo IV avvenne, come la precedente, senza soluzione di continuità. A Tepe Yahyā non è necessario ricorrere alla solita motivazione del cambiamento culturale cioè all'arrivo di un nuovo popolo con l'etichetta di « protoelamita ».

Tuttavia l'architettura subí notevoli trasformazioni. Il sito cessò di avere una funzione residenziale per assumerne una amministrativa. Una vasta struttura che abbiamo portato alla luce al livello IV-C del tumulo è accuratamente orientata in modo che i suoi muri corrano da nord a sud e da est a ovest. I muri sono formati da tre corsi di mattoni a stampo di nuove dimensioni. I mattoni a stampo piú antichi misuravano 15 × 15 × 30 centimetri; i piú recenti 24 × 24 × 12 centimetri, un terzo piú larghi e spessi meno della metà. All'interno di questa grande struttura, che contiene un numero indeterminato di vani, ne sono stati individuati cinque, ma finora si è sgombrata completamente solo parte di uno di essi. La struttura e la stanza parzialmente scavata continuano verso il centro del tumulo e quindi le loro dimensioni sono ancora da definire.

La parte della stanza che è stato possibile esaminare misura 3 × 6 metri. Il contenuto ne denuncia chiaramente la funzione commerciale. Tra gli oggetti rinvenuti nella stanza vi sono ciotole con il bordo svasato fatte di materiale grossolano. Questi recipienti sono simili a quelli ritrovati in numerosi altri siti della Mesopotamia e si ritiene siano serviti come misure di riferimento. Nella stanza furono anche trovate tre grandi giare per provviste che si rivelarono vuote; vicino a esse vi erano 24 suggelli: tappi per giare fatti di argilla con l'impronta di un sigillo. I sigilli impiegati a questo scopo erano di tipo cilindrico; i motivi impressi ricordano quelli dei sigilli cilindrici trovati a Susa, la capitale elamita dell'area mesopotamica. Questo ritrovamento lascia intravedere la possibilità che all'inizio del periodo IV giungessero a Tepe Yahyā merci provenienti da Susa.

Sul pavimento della stanza c'erano 84 tavolette d'argilla nuove e altre sei che recavano delle iscrizioni. Le tavolette hanno tutte le medesima forma e sono fatte di argilla cruda marrone scuro, hanno un profilo convesso e misurano 45 × 51 millimetri. Sulle sei tavolette sono incise complessivamente 17 righe di scrittura protoelamita. Le incisioni sono state impresse da uno stilo sull'argilla molle e si leggono da destra a sinistra lungo l'asse principale della tavoletta e dall'alto in basso. Quando la iscrizione continuava da un lato all'altro della tavoletta, lo scrittore la ruotava intorno al suo asse principale di modo che l'ultima riga della faccia anteriore e la prima di quella posteriore venivano a trovarsi in corrispondenza.

Si sta attualmente provvedendo a decifrare le iscrizioni di Tepe Yahyā. Da un esame preliminare è emerso che si tratta di registrazioni o ricevute di merci. Il fatto che siano state trovate insieme nella stessa stanza tavolette scritte e altre identiche sotto ogni aspetto, ma intatte, è una prova evidente che le incisioni venivano fatte al momento. Pertanto le merci in esse descritte devono essere state di passaggio nell'area amministrativa in entrata o in uscita.

Prima della scoperta di Tepe Yahyā, le uniche altre tavolette protoelamite conosciute erano quelle che provenivano da Susa o da Sialk nell'Iran nordoccidentale. A Susa ne furono trovate circa 1500, a Sialk 19 soltanto. Iscrizioni protoelamite sono state rinvenute recentemente a Shahdāb, un sito a nord di

Kerman, dove sono in corso gli scavi organizzati dall'Iran Archaeological Service. Tali iscrizioni molto brevi non sono incise su tavolette ma su vasellame e raggiungono un massimo di 7 segni.

Nell'unica struttura IV-B esaminata finora si rivela una seconda variazione nello stile architettonico. Si tratta di una costruzione di 2,74 × 2,32 metri, il cui orientamento non tiene conto delle direzioni nord-sud e est-ovest. È costituita di mattoni la cui forma e dimensione è ancora una volta diversa: essi sono essenzialmente oblunghi e la loro lunghezza può essere di 35,5 o 43 centimetri mentre le altre due dimensioni rimangono invariate. La struttura è suddivisa in due stanze principali e in altre piú piccole che contengono recipienti di argilla cruda per le provviste. Le pareti sono spesse quanto un mattone e la loro superficie interna è intonacata.

I vasi delle provviste in una delle stanze principali contenevano ancora varie libbre di frumento. Il frumento era carbonizzato e questo, insieme al fatto che le stuoie sul pavimento e i mattoni del muro erano bruciati, indica che la costruzione era stata distrutta dal fuoco. Tra i detriti sul pavimento furono rinvenuti sigilli cilindrici e, per la prima volta a Tepe Yahyā, anche sigilli a stampo.

Sono stati anche scoperti utensili di bronzo; aghi e ceselli, portati alla luce insieme a manufatti di steatite, venivano probabilmente utilizzati per la lavorazione di quest'ultima. Tylecote e McKerrell hanno scoperto che un pugnale lungo circa 18 centimetri era stato fabbricato forgiando metallo fuso con la stessa tecnica degli utensili del periodo V. Le analisi rivelarono che il pugnale, a differenza dei precedenti manufatti di bronzo all'arsenico, era costituito da una lega contenente il 3 % di stagno. In questa zona dell'Iran non c'è stagno e ciò significa che a Tepe Yahyā devono essere stati importati il pugnale, o lo stagno in esso contenuto o un lingotto di questa lega.

La prova che a Tepe Yahyā si conobbe la scrittura non piú tardi che altrove è una scoperta fondamentale per la preistoria. Seconde per importanza sono forse le numerose testimonianze che indicano la funzione economica veramente unica avuta dalla città a partire dalla fine del quarto millennio a.C..

Le analisi effettuate col carbonio 14 hanno rivelato che la fase IV-B di Tepe Yahyā, iniziata al chiudersi del quarto millennio, arrivò a coprire i primi due secoli del terzo. Durante quel periodo la città fu un importante centro di rifornimento per i manufatti di steatite.

Gli oggetti di steatite, dalle semplici perle alle ciotole decorate, tutte molto simili, sono stati trovati in siti dell'Età del Bronzo lontani tra loro come Mohenjo-Daro, il famoso centro culturale di Harappa sull'Indo, e Mari sull'alto Eufrate a 2400 chilometri di distanza. Tuttavia la Mesopotamia era una regione povera di risorse naturali, compresa la steatite, e anche le genti di Harappa nella valle dell'Indo sembra non potessero rifornirsi localmente di parecchi materiali richiesti. Come si potevano procurare queste sostanze esotiche? Testi sumeri e accadici localizzano le fonti di alcune merci di lusso in luoghi i cui nomi non hanno per noi oggi alcun significato: Dilmun, Maluhha e Magan.

Le ricerche condotte da alcuni danesi sull'isola di Bahrein nel Golfo Persico hanno confermato la supposizione che quest'isola sia proprio l'antica Dilmum. Vi è anche una certa probabilità che la regione o luogo noto come Maluhha si trovasse nella valle dell'Indo. Anche prima che iniziassimo il nostro lavoro a Tepe Yahyā, si pensava che la regione nota come Magan potesse essere localizzata in qualche punto dell'Iran sudorientale. I nostri scavi hanno contribuito a rafforzare notevolmente questa ipotesi. In un testo sumerico piuttosto frammentario si legge: « Possa la terra di Magan (procurarti) molto rame, la forza di... diorite, pietra "u", pietra "shumash"». Questi due nomi di pietre che non sono stati tradotti potrebbero indicare la steatite? Tepe Yahyā e il suo hinterland erano il centro di questi commerci. Proviamo a esaminare le testimonianze offerte dal luogo.

A Tepe Yahyā è stata trovata una quantità di steatite superiore a quella di qualunque altro sito del Medio Oriente. In totale vi sono piú di mille tra frammenti, pezzi non finiti e oggetti intatti, la maggior parte dei quali risale al periodo IV-B. Tra i pezzi intatti troviamo perle, bottoni, sigilli cilindrici, statuette e ciotole. Blocchi di steatite non lavorata, vasi parzialmente scavati e sigilli e perle non finiti dimostrano che Tepe Yahyā era un sito manifatturiero e non un semplice punto di passaggio dei traffici.

Alcune ciotole di steatite sono lisce, altre invece elaboratamente incise. Le

Due sigilli cilindrici trovati a Tepe Yahyā nello strato immediatamente superiore al primo insediamento protoelamita sono qui fotografati a fianco delle impronte che essi lasciano. Le incisioni rappresentano coppie di figure umane con attributi soprannaturali e sono simili a quelle che compaiono sui sigilli di origine mesopotamica.

La piú antica struttura di Tepe Yahyā, attribuita al periodo VI, è un'area adibita a magazzino formata da piccole unità che misurano 1,5 metri di lato. Solo alcune di queste unità sono provviste di porte; evidentemente vi si accedeva attraverso un'apertura ricavata nel tetto. I muri erano costituiti da mattoni di fango essicati al sole, fatti a mano piuttosto che in stampi, o semplicemente di fango impastato. Il cerchio bianco (*a sinistra*) indica il punto in cui è stata rinvenuta una figurina femminile.

Nella fotografia sono visibili le tracce di due costruzioni elamite di Tepe Yahyā. I muri dell'edificio piú antico (*a sinistra*), che risale al 3500 a.C., vennero costruiti con mattoni di fango fatti a stampo che misurano 24 centimetri di lato. I muri corrono da nord a sud e da est a ovest. Quelli dell'edificio piú recente (*a destra*) non sono orientati nello stesso modo; esso fu costruito dopo il 3000 a.C. con mattoni di fango fatti a stampo di forma allungata e di due dimensioni. Sembra che ambedue le strutture fossero adibite a funzioni amministrative piuttosto che residenziali. In quella piú antica sono stati rinvenuti vasi per provviste e ciotole di misurazione. In un angolo sono visibili 84 tavolette per scrivere, ammucchiate l'una sull'altra.

decorazioni sono di tipo geometrico o presentano motivi curvilinei, animali e figurine umane; tra questi motivi ve ne sono alcuni che si ritrovano su numerose ciotole di steatite rinvenute in Mesopotamia e nella valle dell'Indo in siti dell'Età del Bronzo. Inoltre i disegni delle ceramiche provenienti da località come Bampur, a oriente di Tepe Yahyā, e Umm-an-Nai sul Golfo Persico, sono gli stessi di alcune ciotole di steatite dei livelli IV-B.

Durante la nostra stagione di scavi del 1970 riuscimmo a localizzare quella che probabilmente era una delle fonti di steatite di Tepe Yahyā. Un affioramento di roccia delle Montagne Ashin a circa 30 chilometri dal tumulo rivela la presenza di un'antica miniera superficiale. È improbabile che questa sia stata l'unica fonte di rifornimento. I giacimenti di steatite sono spesso associati a quelli di asbesto e cromite. A soli 16 chilometri da Tepe Yahyā c'è una miniera di cromite e in alcune pietre portate alla luce durante lo scavo abbiamo notato vene di asbesto. Una perlustrazione delle montagne settentrionali potrebbe condurre a localizzare altri affioramenti di steatite.

Considerando le grandi quantità di steatite rinvenute in questa località, le prove che molti oggetti venivano fabbricati sul posto, la disponibilità di materiale grezzo nelle vicinanze e la presenza, in Mesopotamia e nella regione di Harappa, di ciotole i cui motivi sono l'esatta ripetizione di quelli trovati a Tepe Yahyā, non si può fare a meno di concludere che questa città è stata un centro di scambi e un importante produttore di tale materiale. Prima di mettere in rilievo la piú vasta portata di questa attività commerciale in un'area cosí remota geograficamente, descriveremo a grandi linee gli altri periodi di occupazione di Tepe Yahyā.

Non vi è quasi piú nulla da aggiungere sulla fase finale del periodo IV che arrivò pressappoco fino al 2200 a.C.. È a questo punto che a Tepe Yahyā si riscontra una soluzione di continuità: la rioccupazione del sito nell'Età del Ferro che durò approssimativamente dal 1000 al 500 a.C. costituisce il periodo III. Ciò è comprovato da una serie di strati con tracce di insediamenti e da un vasellame che rivela strette connessioni di tipologia e di forma con quello prodotto nello stesso periodo nell'Iran nordoccidentale. Non abbiamo ancora portato alla luce nessuna importante struttura del periodo III; sia il tipo di cultura sia i rapporti di Tepe Yahyā con i suoi vicini dell'Età del Ferro rimangono oscuri.

Il periodo II di Tepe Yahyā, che è rappresentato dai 200 e piú anni di occupazione achemenide, fu un'epoca di costruzioni su grande scala. Il materiale impiegato continua ad essere costituito dai mattoni di fango, ma non è ancora stata portata alla luce una struttura completa. Le due ampie stanze scavate finora danno comunque l'impressione che il sito abbia avuto ancora una volta una funzione almeno parzialmente residenziale.

I circa 600 anni successivi di occupazione parta e sassanide, che costuiscono il periodo I, rappresentano lo stadio finale della civiltà urbana di Tepe Yahyā. Abbiamo rinvenuto tracce di architetture su larga scala che comprendevano alcuni cortili e i resti di una massiccia piattaforma di mattoni di fango disposti a quattro corsi l'uno sull'altro. Nel periodo sassanide (all'inizio del terzo secolo) le rovine che si erano venute accumulando per migliaia di anni avevano portato il tumulo a una considerevole altezza; la struttura parzialmente scavata doveva essere una cittadella che si innalzava sulla sua sommità.

Tavolette incise rinvenute a Tepe Yahyā (in fotografia) accanto ai disegni che riproducono i simboli che vi sono scritti. Finora sono state rinvenute 90 tavolette di cui solo 6 incise, tutte della medesima forma. Le iscrizioni, in protoelamita, venivano incise da destra a sinistra per tutta la lunghezza della tavoletta premendo l'estremità acuminata di uno stilo nell'argilla molle.

La maggior parte della ceramica sassanide è costituita da giare, per provviste, spesse e grossolane. Una gran quantità di perle, numerose bottigliette di vetro e ceramica, probabilmente contenitori per profumi, suggeriscono l'idea di una certa prosperità durante il periodo I. Inoltre la presenza di spade di ferro e di bronzo, di asce e di punte di freccia rivela l'esistenza di un ambiente militare. L'unico oggetto artistico, una statuetta d'argilla, rappresenta un guerriero con una particolare acconciatura (*si veda la figura nella pagina seguente in basso*). In seguito, da un certo periodo del quinto secolo in poi, Tepe Yahyā fu occupata solo da intrusi occasionali o da nomadi di passaggio; nessuno dei visitatori si fermò o ebbe modo di innalzare costruzioni di qualche importanza.

Quale funzione ebbe la Tepe Yahyā elamita nella trasmissione della tradizione urbana da occidente a oriente? La posizione della città indica che la cultura elamita, che ora si rivela assai piú diffusa di quanto si credesse un tempo, fu un valido strumento di contatto tra la prima civiltà urbana della Mesapotamia e la civiltà che sbocciò in seguito nella valle dell'Indo. Sembra che gli Elamiti della Persia orientale si siano spinti molto piú in là. Per aver conferma di questa possibilità è necessario cercare la prova di un contatto diretto, prescindendo dal commercio che si svolgeva attraverso gli intermediari, tra la Mesopotamia e la valle dell'Indo.

In alcuni siti della Mesopotamia si è rinvenuto un piccolo numero di manufatti che si suppone, o si è certi, provengano da Harappa. Poiché molti degli scavi archeologici risalgono a un secolo fa, non sorprende che l'età e la localizzazione originale di parecchi di questi manufatti possano essere valutate solo in modo approssimativo. Tuttavia in Mesopotamia si sono trovati sei sigilli a stampa, un sigillo cilindrico e un suggello di argilla, tutti tipici della cultura di Harappa, che provano l'esistenza di un qualche contatto tra le due civiltà. Alcuni sigilli portano iscrizioni della cultura di Harappa, su altri, oltre ai caratteri, vi sono figure di animali il cui stile è quello caratteristico di questa cultura: un unicorno, un elefante, un rinoceronte. Questa è una prova che dovette esserci un contatto; ma fu diretto o indiretto?

L'unico suggello dell'Indo trovato in

Gli oggetti di bronzo risalenti al periodo culminante della lavorazione e del commercio della steatite a Tepe Yahyā comprendono due ceselli (a sinistra) e tre elementi aghiformi; gli oggetti piú lunghi misurano 16 centimetri. Il bronzo non veniva ottenuto con una lega ma direttamente da rame che conteneva allo stato naturale una notevole quantità di arsenico. Questi ricchi minerali grezzi venivano importati.

Le ciotole di steatite decorate sono un esempio dei numerosi oggetti fabbricati a Tepe Yahyā ed esportati a oriente e a occidente durante la prima metà del terzo millennio a.C. Frammenti di ciotole con decorazioni simili a quelle di questi frammenti provenienti da Tepe Yahyā sono stati rinvenuti dalla Mesopotamia alla valle dell'Indo.

Questa statuetta di alabastro, alta circa 25 centimetri, raffigurante un ariete è uno dei numerosi oggetti fatti di materiale importato trovati a Tepe Yahyā e riferentesi al periodo del primo insediamento urbano del 3800 a.C. circa. A conferma delle relazioni commerciali che dovettero esistere tra la città e regioni remote, sono stati rinvenuti, oltre all'alabastro, madreperla del Golfo Persico, marmo, turchese e tormalina.

Statuetta di guerriero che risale alla fine dell'occupazione di Tepe Yahyā, quando un avamposto militare sassanide si insediò sulla sommità del tumulo dal terzo secolo a.C. fino al 400 d.C.. Da allora i nomadi furono i soli visitatori della città morta.

Mesopotamia fu scoperto dall'archeologo francese G. Contenau a Umma, nell'Iraq meridionale, negli anni venti. Ciò suggerisce l'ipotesi dell'arrivo in questo luogo, dal territorio di Harappa, di merci che venivano identificate mediante il marchio personale del mittente, impresso prima della spedizione. I sette sigilli invece si prestano a interpretazioni molto piú equivoche. I contatti mesopotamici con l'Indo non avevano nessun punto in comune con i successivi scambi commerciali tra la Mesopotamia e, per esempio, l'impero ittita a occidente. In quel caso i commercianti assiri venivano alloggiati in quartieri speciali di roccaforti ittite come Kültepe e Hattusha.

In breve, non vi sono prove sicure che i mesopotamici abbiano visitato la valle dell'Indo per risiedervi o per stabilire contatti commerciali o che le genti di Harappa abbiano fatto l'inverso.

Per qual motivo dunque i sigilli dei commercianti di Harappa si trovavano in Mesopotamia? A che cosa servivano i tre sigilli rinvenuti a Ur, i due di Kish e i due di Tell Asmar? Finora non è stato possibile dare una risposta convincente a tali interrogativi. Si è tentati di considerare questi sigilli, tutti oggetti di pregevole fattura, non come credenziali ma come « ricordo » di contatti commerciali indiretti. Al tempo stesso si pone un altro interrogativo sconcertante. In Mesopotamia sono stati trovati alcuni oggetti provenienti dalla valle dell'Indo. Per quale motivo in questa zona non è stato rinvenuto alcun manufatto mesopotamico?

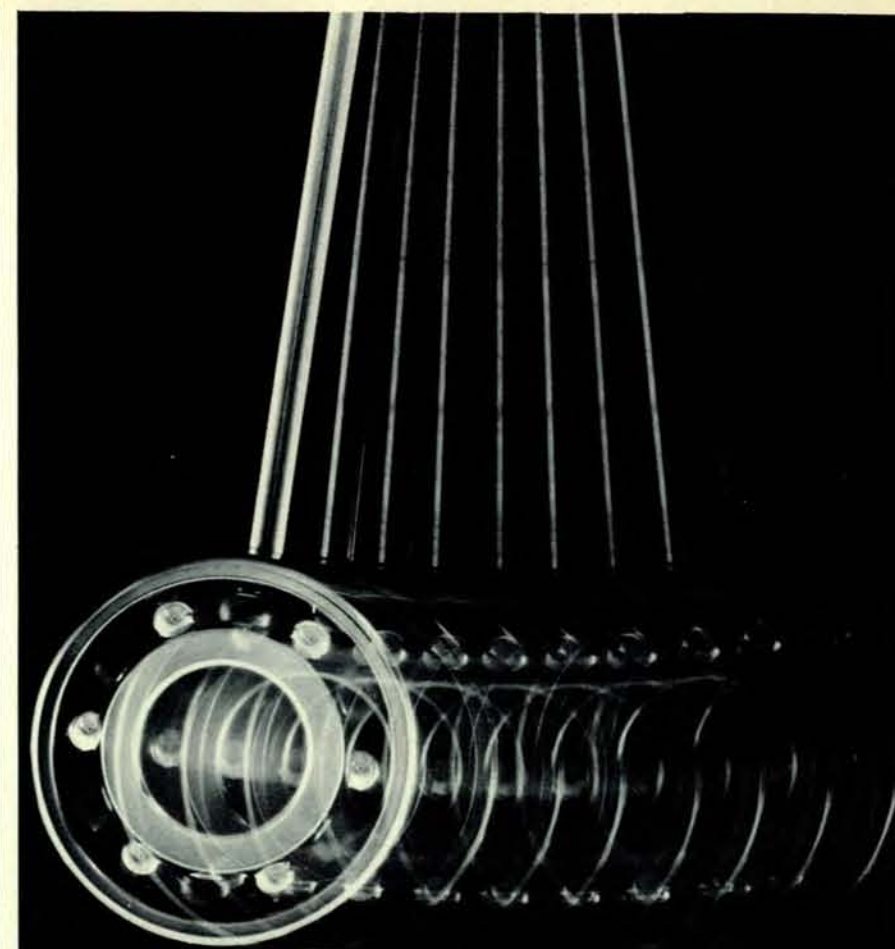
Mancano prove sicure di un contatto commerciale diretto fra queste due civiltà. Sono però noti altri generi di commercio. Tra i piú antichi e i piú diffusi è il semplice scambio che può comportare, in modo piú o meno deliberato, un numero qualsiasi di intermediari. L'interpretazione di questo tipo di scambio è estremamente difficile per l'archeologo: stile e funzione, di per sé intangibili, possono diffondersi insieme alle merci.

Un sistema di scambi basato su un solo intermediario sembra fornire il modello teorico che piú si avvicina alla situazione di Tepe Yahyā. Tale sistema è noto come commercio « a smistamento centrale »; a nostro avviso, durante il periodo elamita, Tepe Yahyā rivestiva proprio questa funzione nella Persia sudorientale.

Un centro di questo genere può trovarsi al di fuori della sfera d'influenza dell'una o dell'altra comunità principale e, contemporaneamente, produrre merci o controllare risorse naturali a cui aspirano entrambe. Oltre a esportare i propri prodotti, esso può trasportare le merci fabbricate dall'una o dall'altra comunità e questa può anche essere la sua unica attività. Bahrein – l'antico Dilmun – costituisce un valido esempio di centro la cui prosperità dipendeva dallo smistamento di merci dirette in Mesopotamia. Indipendentemente dall'importanza che queste attività potevano rivestire per Tepe Yahyā, è indubbio che la funzione principale della città in epoca elamita era quella di fabbricare ed esportare oggetti di steatite.

I nomi dei siti mesopotamici in cui sono state rinvenute ciotole di steatite identiche per forma e motivi decorativi a quelle trovate a Tepe Yahyā costituiscono un elenco di riferimento per l'archeologo: Adab, Mari, Tell Asmar, Tell Aqrab, Khafaje, Nippur, Telloh, Kish, Al 'Ubaid e Ur. Ciotole nello stile di Tepe Yahyā sono state rinvenute a Mohenjo-Daro sull'Indo e a Kulli-Damb nel Belucistan pakistano. Oltre alle ciotole svasate del tipo di Uruk, testimonianza dei contatti di Tepe Yahyā con l'occidente, nel tumulo sono state trovate ceramiche Nay, un tipo di vasellame dipinto, originario della valle dell'Indo e anteriore al sorgere della civiltà di Harappa, che testimonia i contatti della città con l'oriente.

Tuttavia Tepe Yahyā non fu il solo centro di smistamento commerciale della Persia orientale. Sembra piuttosto che sia stata una delle numerose città comprese in una federazione elamita a carattere locale situata in posizione intermedia tra i territori delle due civiltà. Sembra che Shahr-i-Sokhta, un sito a 400 chilometri a nord-est di Tepe Yahyā, sia stato un altro centro impegnato nell'esportazione dell'alabastro locale e nello smistamento di lapislazzuli provenienti dall'Agfanistan. Si è già parlato dei legami tra Tepe Yahyā e altri possibili centri di smistamento della regione — come Tal-i-Iblis, Shahdāb e Bampur — legami di cui sono testimonianza soprattutto le somiglianze rilevate nella ceramica.

Come poté questo remoto dominio elamita, che nel caso di Tepe Yahyā precedette di almeno tre secoli il sorgere della civiltà di Harappa, influire sullo sviluppo della valle dell'Indo? Nonostante le nuove prove dell'esistenza di reti commerciali fin dall'inizio del Neolitico, vi è una diffusa tendenza a considerare il commercio solo come un sottoprodotto dell'urbanesimo. Ma il commercio è stato certamente uno degli stimoli principali che hanno portato alla civiltà urbana. È proprio quanto avvenne, a nostro avviso, nell'antico Kerman e nel Belucistan persiano.

Un gozzo di modeste proporzioni è evidente in questo particolare del ritratto di Maria de' Medici, dipinto nel 1625 da Rubens e oggi al Prado di Madrid. Nel tardo Rinascimento un gozzo moderato era considerato un particolare pregio estetico.

Il gozzo endemico

La malattia, conosciuta fin dalle prime epoche storiche è attualmente un'afflizione caratteristica della povera gente. Spesso infatti una dieta non equilibrata non compensa la deficienza iodica di parecchi terreni

di R. Bruce Gillie

Il « collo rotondo e regolare », che Rubens attribuí a Maria de' Medici (*si veda l'illustrazione nella pagina a fronte*) non è altro che un gozzo (o struma), ipertrofia di compensazione della tiroide. La tiroide è un roseo cuscinetto di tessuto avvolto in parte attorno alla trachea e all'esofago; è una ghiandola priva di dotto escretore (endocrina), presente nei vertebrati, che versa direttamente nel sangue gli ormoni intesi a regolare il tasso di sviluppo dell'individuo e il metabolismo. Il gozzo è quindi la manifestazione insolitamente evidente di un disordine endocrino, che fin dai primissimi tempi della storia umana, proprio a causa della sua evidenza, suscita di volta in volta stupefazione e timore.

Il gozzo ha cause varie e diverse: una malattia, certi difetti di sviluppo, le condizioni ambientali. Il gozzo endemico, cosí definito in quanto colpisce in misura rilevante una data popolazione, è quasi sempre conseguenza di una deficienza dietetica di iodio, substrato indispensabile alla sintesi degli ormoni tiroidei tiroxina e triodotironina. Oggi il gozzo dovuto a carenza di iodio si può agevolmente prevenire o curare con l'assunzione di dosi minime di questo elemento, ma per secoli è stato uno dei morbi piú persistenti e diffusi che abbiano mai afflitto l'umanità: ancora nel 1960 ne soffrivano non meno di duecento milioni di persone.

La secrezione degli ormoni tiroidei è il passaggio obbligato di uno dei sistemi di retroazione piú delicatamente equilibrati che regolino l'ambiente interno dei vertebrati. Gli impulsi provenienti dal sistema nervoso inducono l'ipotalamo, situato alla base del cervello, a liberare una neurosecrezione, il fattore tirotropo (TRF, thyrotropin-releasing factor), che si versa nei vasi portali direttamente adducenti all'ipofisi, quella ghiandola maestra grossa come un pisello che regola l'attività della tiroide e di altre ghiandole endocrine. Sotto lo stimolo del fattore tirotropo, la pituitaria (o ipofisi) versa nel sangue la tirotropina, o ormone tireotropo (TSH, thyroid-stimulating hormone), che a sua volta costringe la tiroide a sintetizzare e a secernere i propri ormoni. Si tratta di un sistema ad autoregolazione: l'eventuale eccesso di ormoni tiroidei nel sangue blocca l'attività dell'ipotalamo e dell'ipofisi, riducendo la secrezione dell'ormone tireotropo; quando invece la concentrazione di ormoni tiroidei è carente, l'ipofisi reagisce, secernendo una maggior quantità di ormone tireotropo e ristabilendo cosí il normale quantitativo di ormoni tiroidei (*si veda l'illustrazione a pag. 64*).

Se la tiroide è sana e se nel sangue vi sono ioduri (iodio ionico) a sufficienza, l'ormone tireotropo stimola la cattura dello iodio da parte della tiroide, favorendo inoltre la sintesi della tiroxina e della triodotironina nell'ambito dei follicoli tiroidei (*si veda l'illustrazione a pag. 66*). In assenza di una quantità sufficiente di ioduri si blocca la sintesi della tiroxina: accade cosí che l'afflusso dell'ormone tireotropo, una volta venuta a mancare la necessaria ritenzione, provoca l'aumento quantitativo e la modifica morfologica delle cellule costituenti i follicoli tiroidei i quali, a tempo debito, si distendono. Questa proliferazione cellulare compensatoria e la conseguente distensione dei follicoli, che costituiscono appunto il gozzo, possono riportare a un livello utile alla vita normale la produzione degli ormoni tiroidei.

Il primo riferimento storico al gozzo che si conosca è un documento cinese risalente all'incirca al 3000 a.C. È interessante notare come lo scritto non descriva soltanto i sintomi del male, ma raccomandi addirittura una cura efficace: l'ingestione di alghe marine e di spugna arrostita, materiali che contengono infatti un buon quantitativo di iodio.

I papiri egizi di Ebers, risalenti al 1500 a.C. circa, descrivono due possibili terapie per il gozzo: l'asportazione chirurgica della ghiandola (che, se mai venne applicata, dovette essere un procedimento rischiosissimo) e l'ingestione di sale (presumibilmente contenente iodio) di una data zona del basso Egitto.

Ippocrate attribuiva il gozzo all'acqua potabile di certi luoghi. Giovenale, Vitruvio e Giulio Cesare restarono impressionati dalle insolite proporzioni del collo degli abitanti di certe regioni alpine. Cesare era anzi convinto che il collo grosso fosse caratteristica nazionale dei celti. La parola « gozzo », tra l'altro, deriva dal latino *guttur*, gola.

I medici dell'antica Roma notarono che persino nell'individuo normale la dimensione della tiroide può variare in periodi di sollecitazione fisiologica particolare, come durante la pubertà e in occasione delle mestruazioni o della gravidanza nelle donne. In particolare rilevarono che le circostanze emotive e fisiche attinenti all'attività sessuale iniziale delle giovani spose potevano provocare appunto l'ingrossamento della ghiandola. Da qui ebbe origine il rito antico di misurare il collo della sposa prima e dopo la prima settimana di matrimonio con un nastro cerimoniale: se la circonferenza risultava aumentata, le nozze si dovevano considerare consumate.

Il gozzo di proporzioni modeste, essendo perfettamente compatibile con la vita normale, in quanto non provoca né dolori né compromissioni particolari, non era necessariamente riguardato come un fatto morboso, tanto che presso talune culture veniva accolto quale segno del favore divino e, presso altre, quale distintivo di beltà. In Europa fu spesso attribuito a certe gravi trasgressioni del codice sociale o religioso, per

Nel nostro paese il gozzo endemico è presente con una diffusione che tocca praticamente tutte le regioni. Stime non recentissime, ma tuttora utilizzabili come elementi orientativi, indicano che il 5-10 % della popolazione italiana (cioè un numero di persone intorno ai 2,2-5 milioni) è affetta da questa condizione morbosa. Per una parte dei casi si tratta di un fenomeno transitorio, essendo noto che la frequenza del gozzo varia in funzione dell'età e del sesso, per cui si osserva talora nel maschio adolescente o adulto la scomparsa del gozzo insorto in periodo puberale. La geografia del gozzo copre, con frequenza variabile della condizione morbosa, tutto l'arco alpino dalla Liguria alla Venezia Giulia.
In questo settore possono ricordarsi alcuni tra i focolai piú conosciuti e meglio descritti: Cuneo, Saluzzo (città in cui si verificò, come del resto in altre zone del Piemonte, una recrudescenza di forme acute di gozzo durante la seconda guerra mondiale); val d'Aosta, Sondrio e la Valtellina, le vallate bergamasche e bresciane; alcune zone del Trentino-Alto Adige, la Valsugana, i dintorni di Belluno e Udine. Nei focolai delle zone prealpine ne risultano comunemente affetti il 20-30 % dei ragazzi in età scolare, mentre nelle zone di montagna la frequenza è sovente superiore, raggiungendo il 50-60 % e, in alcuni casi, anche valori piú elevati. Assai meno frequente, anche se non assente, è il gozzo nella pianura padana, mentre esso è nuovamente rappresentato, con frequenze non dissimili da quelle citate per le zone prealpine e alpine, nell'Appennino modenese, toscano, marchigiano, umbro, nell'ascolano e nel viterbese. Altre aree di endemia sono state segnalate anche piú a sud, nella provincia di Potenza, nella penisola sorrentina, nel Gargano e in provincia di Catanzaro. In Sicilia sono stati descritti focolai a Nicosia (Enna) e in provincia di Messina e in Sardegna in provincia di Cagliari e nella parte centro-occidentale dell'isola (Santulussurgiu).
Non sempre si è potuta rilevare, negli studi condotti nelle regioni italiane, una relazione tra deficit iodico alimentare e frequenza di gozzo e sono stati di volta in volta invocati altri fattori causali (le sostanze gozzigene menzionate da Bruce Gillie, l'elevato contenuto calcareo o siliceo delle acque alimentari) o puramente ipotetici. Indipendentemente dalle cause, la presenza del gozzo è *sempre* l'indice visibile di uno squilibrio patologico della ghiandola tiroidea che può evolvere in modi diversi: a) regredendo verso la normalità con scomparsa del gozzo; b) (evenienza piú frequente) mantenendosi funzionalmente compensato con permanenza del gozzo, in forma diffusa o nodulare, con la possibilità, oltre all'inevitabile danno estetico e psicologico, di disturbi meccanici da compressione che rendono necessaria la rimozione chirurgica; c) dando luogo a sequele quali l'ipotiroidismo, di solito lieve, o l'ipertiroidismo; d) creando inoltre alcune delle premesse che favoriscono l'insorgere del cretinismo nei discendenti degli affetti di gozzo. La somministrazione sistematica di iodio, nella misura di 100-200 microgrammi (milionesimi di grammo) al giorno, è in grado di ovviare quasi completamente a questi inconvenienti in quanto riduce drasticamente, fino a valori prossimi a zero, la frequenza di comparsa del gozzo. L'utilità pratica di questa constatazione non è minimamente sminuita dal fatto che non è ancora del tutto chiaro perchè lo iodio riesca a svolgere questa netta azione preventiva: se agendo come puro sostitutivo, là dove esista un deficit iodico alimentare o, in altri casi, intervenendo in modo piú complesso nella catena di reazioni fisiologiche che si traducono alla fine nella comparsa del gozzo. L'esperienza pluriennale di un paese come la Svizzera, in cui la profilassi iodica è stata realizzata semplicemente obbligando la vendita di sale da cucina iodato in misura di 5 o 10 mg di iodio per kg, ha definito chiaramente l'efficacia e la parallela innocuità di questo tipo di profilassi. Nel nostro paese la situazione, in questo settore della salute pubblica appare alquanto sconfortante: scriveva qualche anno fà un autorevole studioso italiano del gozzismo (Costa e coll., 1966) « ...nella maggior parte delle località di endemia del nostro paese la profilassi iodica o non ha potuto iniziare o, una volta iniziata, è miseramente fallita ». Le ragioni di questo fallimento sono di varia natura, sia tecniche che amministrative. Ma, fondamentalmente, si possono ricondurre alla mancata adozione di provvedimenti che rendano *obbligatorio* il consumo di sale iodato. In sede di Organizzazione Mondiale della Sanità è stato fatto notare che il solo provvedimento capace di promuovere un adeguato consumo di sale iodato è l'introduzione della profilassi iodica obbligatoria. Malgrado questo ancora oggi molti paesi, tra cui l'Italia, persistono a non percorrere la via della prevenzione obbligatoria. Di recente una commissione di lavoro costituita presso il Ministero della Sanità ha disposto di rilanciare la propaganda della profilassi iodica e di facilitare la distribuzione del sale iodato. Ma è dubbio che questa lodevole iniziativa basti da sola per attaccare alla base il problema, rilevante nel nostro paese, del gozzismo endemico. *R.S.*

esempio ai trafugamenti di oggetti preziosi dalle tombe dei santi. Nella Germania medievale si pensava che la condizione patologica derivasse da fatiche estenuanti, comprese quelle relative al parto. E questa è appunto la spiegazione razionale della costumanza, oggi dimenticata, di legare un cordoncino attorno al collo delle partorienti.

Quanto alle cure, non era tanto facile trovarne. Il procedimento rimasto in auge per secoli consisteva nel perforare la tiroide con un ago dalla punta arroventata. È da presumere che l'intervento provocasse un'infiammazione e che la conseguente fibrosi contribuisse a far scemare la grossezza della ghiandola. È impossibile comunque che prima della fine del secolo scorso la chirurgia vera e propria conducesse a risultati efficaci, se non altro perché una eventuale incisione della tiroide, cosí riccamente alimentata dai vasi sanguigni, avrebbe causato, in quei tempi di chirurgia primitiva, un'emorragia abbondante e inarrestabile. Persino dopo l'avvento delle tecniche d'intervento piú idonee, almeno prima della scoperta delle ghiandole paratiroidi, l'asportazione chirurgica doveva essere pericolosissima: le paratiroidi, ghiandole di piccolissime dimensioni, situate sulla faccia dei lobi laterali della tiroide, regolano la percentuale di calcio nel sangue, per cui la loro rimozione accidentale insieme con la tiroide gozzuta costituisce una minaccia immediata alla vita del paziente.

Va da sé che il vero mistero che circondava il fenomeno del gozzo risiedeva nella funzione della tiroide negli individui sani. Gli antichi anatomisti, impressionati dal grande afflusso di sangue alla ghiandola e incapaci di spiegare secondo i loro criteri la mancanza di un dotto escretore (come è il caso di tutte le ghiandole endocrine), ritenevano che la tiroide non potesse avere alcuna funzione escretiva. Nel medioevo non mancarono gli studiosi di anatomia convinti che questa fosse la sede dell'anima. Altri rivelarono un maggior senso pratico. Nel 1600 l'anatomista italiano Giulio Casserio annotava come la natura gentile avesse specialmente abbellito il sesso femminile con ogni sorta di ornamenti, tra i quali il soprattutto il collo regolare e tornito.

Nel 1656 l'anatomista inglese Thomas Wharton dette una descrizione completa della tiroide, dandole per la prima volta il nome ancor oggi usato, dal termine greco che indicava un grande scudo oblungo, il *thyreos*. Wharton, d'accordo con Casserio, asseriva che la ghiandola servisse ad abbellire il collo (« soprattutto nelle donne, alle quali per tal motivo è stata assegnata una ghiandola piú grande »), ma proponeva che potesse servire anche a tener calda la cartilagine tracheale, « di natura piuttosto fredda », e a lubrificare la laringe, rendendo piú melodiosa la voce. Altri studiosi ritenevano che la tiroide avesse la funzione di allontanare il sangue dal cervello, in modo da tutelare questo importante organo da improvvise variazioni di pressione sanguigna, oppure che si trattasse di un cuscinetto destinato a supportare e a proteggere le strutture della laringe.

Fu solo nel 1895, dopo che i chirurghi ebbero constatato gli effetti dell'asportazione della tiroide e dopo il primo tentativo di terapia con estratto tiroideo, che il medico tedesco Adolf Magnus-Levy dimostrò come la tiroide regolava il tasso del metabolismo basale, ossia la velocità alla quale le cellule dell'organismo consumano ossigeno, cioè la velocità alla quale trasformano in energia vitale le sostanze nutritive. In quello stesso anno il biochimico tedesco Eugen Baumann scoprí per puro caso che la tiroide è particolarmente ricca di iodio. Baumann stava tentando di analizzare il contenuto proteico del tessuto tiroideo, e il suo procedimento abituale consisteva nel far precipitare le proteine da un estratto a mezzo di acido solforico. Un giorno, allungando la mano per prendere il flacone dell'acido solforico che si trovava su di uno scaffale accanto al banco d'analisi, prese per errore un flacone diverso e, prima di rendersi conto dello sbaglio, aveva già aggiunto all'estratto un po' di acido nitrico. Quale non fu la sua meraviglia nel constatare che dal preparato cominciavano a svilupparsi i caratteristici fumi bruno-rossastri del gas iodio. Fu cosí che Baumann si impegnò nello studio e nella descrizione del ruolo dello iodio nella fisiologia della tiroide. Nel 1914, per la prima volta, Edward C. Kendall della Mayo Foundation riuscí a cristallizzare un po' di ormone tiroideo, impresa quanto mai ardua: i trentasette grammi di ormone cristallizzato ottenuti finalmente dallo sperimentatore derivavano da ben tre tonnellate e un quarto di tiroide suina! Nel 1927 Charles Robert Harington della Facoltà di clinica medica del londinese University College e George Barger dell'Università di Edinburgo fissarono una volta per sempre la struttura della tiroxina, confermando cosí le osservazioni di Baumann. Nel frattempo l'opera del chimico tedesco aveva condotto, da un lato, all'interpretazione del gozzo endemico come conseguenza della carenza ambientale di iodio, dall'altro alla semplice ed efficace iodoterapia.

Non appena i ricercatori ebbero cominciato a esplorare l'ecologia delle popolazioni gozzute, accertarono una stretta correlazione tra il gozzo e la possibilità di accesso al mare e, quindi, a una dieta alimentare ricca di iodio. Una cartina appositamente approntata dall'Organizzazione Mondiale della Sanità mostra chiaramente come la forma endemica del gozzo si sviluppi tipicamente nelle zone d'entroterra, soprattutto montagnose (*si veda l'illustrazione in queste due pagine*). Regioni di particolare diffusione del gozzo sono le Alpi, i Pirenei, l'Himalaya e le Ande, ma un'incidenza piuttosto elevata si riscontra altresí nelle pianure interne dell'Italia, del Congo e del bacino nordamericano dei Grandi Laghi.

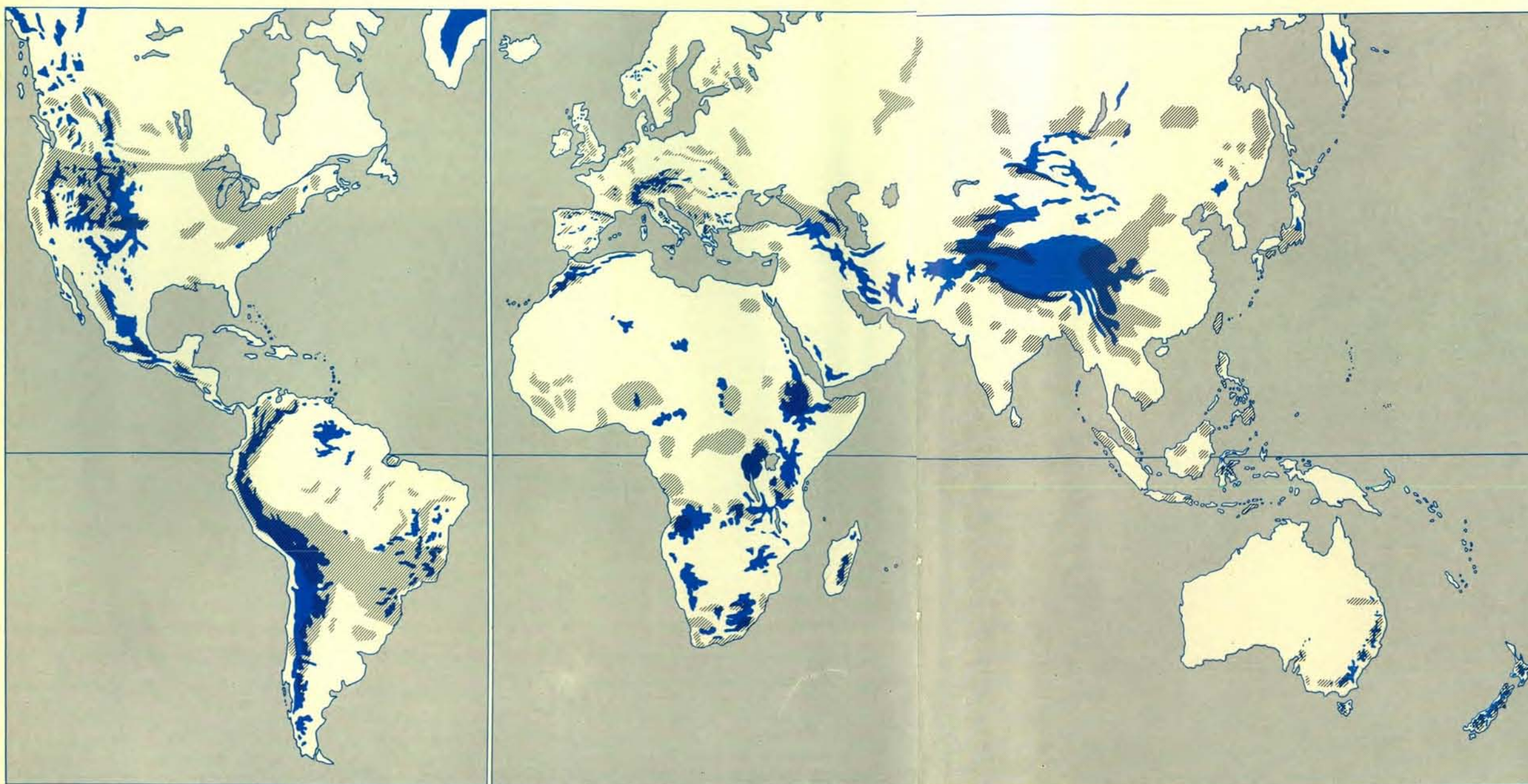
Tuttavia la geografia del gozzo non è affatto semplice: mentre diverse regioni montagnose e d'entroterra non sono abitate da popolazioni gozzute, ci sono aree costiere che imprevedibilmente denunciano una forte presenza di gozzi. Si giunge cosí alla conclusione che elemento ben piú strettamente correlato con l'incidenza del gozzo endemico, rispetto alla pura e semplice distanza dal mare, è il tenore di iodio del suolo fertile. Finché il contenuto di iodio nello strato superficiale del terreno è adeguato, lo iodio consumato nei prodotti agricoli locali è sufficiente a prevenire l'insorgere dell'affezione (bastano 100-200 microgrammi giornalieri a persona). Ebbene, per quanto il tenore di iodio nel suolo fertile sia generalmente piú elevato nelle regioni costiere che non in quelle dell'entroterra, la presenza del prezioso elemento nel suolo è determinata da fattori che poco hanno a che fare con la distanza dal mare. I terreni piú seriamente carenti di iodio sono quelli che hanno subito le glaciazioni piú intense. Il fenomeno della glaciazione è intervenuto in due modi: fratturando rocce ignee vergini, mai prima esposte allo iodio atmosferico, si è lasciato alle spalle immense quantità di nuovo terreno superficiale povero di iodio; lisciviando il suolo originario, lo ha privato dei sali di iodio solubili.

La lisciviazione può anche determinare l'impoverimento di iodio nei terreni agricoli costeggianti i fiumi soggetti a straripamento periodico. Un interessante esempio di questo fenomeno si è osservato nel corso di uno studio condotto in due villaggi situati sulle opposte sponde del fiume Congo, in una zona in cui le intense precipitazioni e le inondazioni periodiche avevano comportato la diminuzione del tenore di iodio nel suolo. La popolazione del villaggio situato sulla riva destra presentava l'80 % di gozzuti, mentre sul terreno alluvionale della riva opposta il gozzo era quasi completamente sconosciuto. Accadeva infatti che lo iodio lisciviato dalle forti piogge nei terreni a monte, veniva trascinato piú a valle, andandosi a depositare sul suolo alluvionale attorno al secondo villaggio, dove si costituiva in concentrazione sufficiente a prevenire l'affezione endemica.

La costante reintegrazione di iodio atmosferico negli strati superficiali del suolo agisce, col tempo, nel senso di un'inversione degli effetti della glaciazione. L'entità della reintegrazione risente in modo vario e complesso della distanza dal mare, delle condizioni di vento prevalenti nella zona e del tenore di iodio caratteristico delle precipitazioni locali. Alla determinazione del contenuto in iodio presente nel suolo concorre anche, in determinati casi, l'accumulo in certe regioni di iodio terrestre naturale. In altre parole anche l'ecologia locale, oltre alla dieta alimentare di base, costituisce elemento rilevante dell'etiologia del gozzo endemico.

La prima testimonianza diretta della carenza di iodio quale causa determinante dell'ipertrofia tiroidea endemica si ebbe a conclusione dell'ampia sperimentazione condotta su cavie di laboratorio, a partire dal 1907, da David Marine e da O. P. Kimball della Facoltà di medicina della Western Reserve University. In seguito questi stessi ricercatori attuarono ad Akron, nell'Ohio, il primo programma su vasta scala di profilassi del gozzo. L'intera operazione, conclusasi nel 1920, interessò 4500 studentesse in età scolare compresa tra l'ultima classe delle elementari e l'ultima delle medie superiori. Mentre una metà delle ragazze serviva da gruppo di controllo, all'altra metà vennero somministrati due grammi di sale iodurato due volte l'anno: al termine di un periodo di due anni e mezzo il 65,4 % del gruppo di trattamento rivelò la regressione del gozzo, mentre furono soltanto cinque le ragazze di questo gruppo nelle quali si poté riscontrare l'ingrossamento della tiroide. Per contro solo il 13,8 % delle studentesse appartenenti al gruppo di controllo rivelarono il regresso della disfunzione, mentre in ben 495 ragazze non trattate si era verificata l'ipertrofia tiroidea.

I risultati della ricerca erano stati appena pubblicati, che già molti privati e molti gruppi di igienisti avevano iniziato il consumo di iodio a dosi massicce, sfiorando in molti casi la soglia del feticismo. Alcuni giunsero al punto di portarsi appresso, appeso al collo, un flaconcino di iodio al quale attingevano un sorso ogni tanto. Lo iodio divenne l'ingrediente magico delle panacee di molti ciarlatani. Ma ecco che, tra lo stupore generale, ci si accorse che molto spesso lo iodio, lungi dal prevenire il gozzo, sembrava quasi stimolarlo. In seguito questo effetto apparentemente paradossale dell'elemento sull'etiologia del gozzo venne spiegato razionalmente da Jan Wolff e da Israel L. Chaikoff della Facoltà di medicina della Università di California a Berkeley, i quali accertarono che proprio l'elevata concentrazione iodica nel plasma sanguigno inibisce la biosintesi e la secrezione dell'ormone tiroideo.

L'idea che a monte delle manifestazioni del gozzo dovesse ricercarsi la responsabilità di qualche agente elementare presente nei cibi o nelle bevande aveva preceduto nel tempo persino la scoperta, quale causa determinante, della carenza di iodio, ma soltanto nel

Le regioni di distribuzione del gozzo endemico e i terreni ai quali il fenomeno si trova spesso associato sono stati cartografati dall'Organizzazione Mondiale della Sanità. Nelle mappe presentate in queste due pagine, le zone in cui il gozzo da carenza di iodio è endemico sono indicate dal tratteggio nero. È evidente come risultino poco colpite le popolazioni del litorale marino, grazie al tenore di iodio di cui sono ricchi gli alimenti del mare. Le zone d'entroterra sono interessate variamente, ma le piú vulnerabili sono quelle caratterizzate dalla particolare costituzione geologica e dall'isolamento dei terreni montagnosi (*in colore*).

1941 si giunse alla scoperta di una sostanza gozzigena. Curt P. Richter e Kathryn H. Clisby dell'Istituto di medicina della Johns Hopkins University, i quali studiavano gli effetti di certi veleni per i topi, si accorsero con viva meraviglia che nei ratti di laboratorio trattati con il preparato tiourea, non sopravveniva la morte, ma la tiroide si ingrossava gradatamente fino a svilupparsi in gozzo. All'incirca nella stessa epoca, in un altro laboratorio della medesima università, Julia B. e Cosmo G. MacKenzie stavano effettuando le loro ricerche sull'incidenza di un nuovo medicamento sulfamidico sulla flora batterica dell'intestino del ratto. Ebbene, anche i MacKenzie osservarono il fenomeno già osservato dai colleghi: la tiroide dei ratti da essi trattati cominciò a manifestare i segni dell'ipertrofia, come se i piccoli animali fossero stati tenuti per settimane e settimane a una dieta iodocarenziale. Evidentemente tanto l'uno quanto l'altro medicamento impedivano all'organismo di utilizzare convenientemente lo iodio che pure si trovava in concentrazioni normali nella dieta alimentare somministrata alle cavie. E da allora in poi sono stati scoperti molti altri composti antitiroidei.

Sul piano teorico questi preparati possono intervenire secondo tre diversi meccanismi d'azione. In primo luogo potrebbero agire a livello intestinale, nel senso di una chelazione (o sequestro) dello iodio, impedendone cosí il normale assorbimento nel circolo sanguigno. In secondo luogo potrebbero agire alla superficie della cellula epiteliale tiroidea, inibendo l'assorbimento selettivo dello iodio contenuto nel sangue che attraversa la ghiandola. (Questa cattura degli ioni di iodio è un processo straordinariamente efficiente: la tiroide riesce a concentrare gli ioni a un livello centinaia di volte superiore a quello della concentrazione ionica nel sangue). In terzo luogo i composti strumogeni potrebbero conquistarsi l'accesso alle cellule tiroidee, sbarrando in una fase cruciale il cammino della biosintesi oppure impedendo la liberazione della tiroxina dalla forma di deposito e attesa in cui si trova nei follicoli tiroidei. Tutto lascia pensare che i meccanismi decisivi siano gli ultimi due: i tiocianati e i perclorati inibiscono i processi di trasporto attivo degli ioduri catturati, il tiouracile blocca la trasformazione degli ioduri in iodio per ossidazione da parte di certi enzimi perossidasici, mentre le solfammidi interferiscono con l'incorporazione della tirosina (*si veda l'illustrazione a pagina 66*).

Il sistema di retroazione negativa che controlla la produzione degli ormoni tiroidei inizia con la neurosecrezione da parte dell'ipotalamo (*1*) del fattore tirotropinogeno (TRF) che, passando direttamente nella pituitaria (*2*), provoca l'immissione di tirotropina, o ormone tireotropo (TSH), nel circolo sanguigno (*3*). Nella ghiandola tiroidea il TSH opera la sintesi e la immissione in circolo degli ormoni tiroidei tiroxina e triiodotironina (*4*); la quantità di ormoni tiroidei che raggiunge l'ipotalamo regola a sua volta la secrezione di TSH, completando cosí il circuito di retroazione negativa. In assenza di iodio, substrato essenziale degli ormoni tiroidei, la produzione ormonica è insufficiente (*5*) a bloccare in tempo il sistema: il conseguente eccesso escretivo di TRF (*6*) e di TSH (*7*) stimola la crescita del tessuto tiroideo povero di iodio (*8*). Un normale follicolo tiroideo, in cui avvengono la sintesi e l'immagazzinamento degli ormoni, si compone di un inviluppo di cellule contenente un colloide, la tiroglobulina (*a*). In assenza di iodio il TSH fa sí che le cellule proliferino e divengano piú colonnari (*b*), e quindi producano altro colloide, finché i follicoli si distendono (*c*) a formare il gozzo cosí come tipicamente si manifesta.

Subito dopo la scoperta di questi composti gozzigeni ci si accorse che il gozzo endemico in certe zone era la conseguenza di analoghi composti che si trovavano in natura nell'alimentazione locale. Tra i generi alimentari contenenti quantità rilevanti di composti strumogeni naturali sono da ricordare la soia e gli appartenenti al genere *Brassica*, comprendente i cavolini di Bruxelles, i cavoli, le rape e altri vegetali commestibili che, assolutamente innocui in una dieta variata, costituiscono un pericolo serio presso quelle comunità che se ne nutrono in modo uniforme.

L'effetto strumogeno della *Brassica* venne dimostrato non molto tempo fa in Tasmania, al largo della costa meridionale australiana. Constatato che l'isola era una zona di gozzo endemico, venne istituito nel 1949 un programma di somministrazione di compresse iodurate agli scolari di età inferiore ai sedici anni. Cinque anni dopo, però, un'indagine di controllo rivelò che l'incidenza del gozzo era aumentata, invece di diminuire. I ricercatori controllarono i dati disponibili, sottoposero a ulteriore verifica la metodologia adottata, ma il risultato si dimostrò esatto. A questo punto F.W.A. Clements e J.W. Wishart, iniziatori del programma, avanzarono l'ipotesi che all'origine del fenomeno dovesse ricercarsi qualcosa di diverso che non la carenza di iodio. Era capitato che nel 1950 il governo australiano avesse iniziato una campagna per la distribuzione gratuita del latte nelle scuole e che l'accresciuta domanda di latte avesse costretto gli allevatori locali a tenere le vacche al pascolo anche in periodi di scarsità stagionale di erba: di conseguenza gli animali si nutrivano di cavolorapa, vegetale ben piú resistente dell'erba al gelo e rigoglioso per tutto l'anno. Il fatto è che il cavolorapa, appartenente al genere *Brassica*, contiene una quantità rilevante di composti strumogeni. Un opportuno supplemento d'indagine rivelò che questi composti, trasmettendosi inalterati al latte vaccino, bloccavano nei consumatori di latte l'utilizzazione dello iodio somministrato.

La biosintesi degli ormoni tiroidei dipende dalla presenza di iodio ionico, soprattutto come ioduro di sodio. Lo ioduro si ossida a iodio elemento e si combina con l'amminoacido tirosina per formare monoiodotirosina e diiodotirosina. Due molecole di diiodotirosina si possono combinare, formando tiroxina; oppure si possono combinare tra loro molecole di mono e diiodotirosina per formare triiodotironina. (In figura si vedono solo i prodotti ormonali, non sono invece indicati, per semplificare lo schema, i sottoprodotti).

Evidentemente esistono dei pericoli intrinseci alla somministrazione di preparati farmaceutici contenenti notevoli quantità di composti potenzialmente strumogeni o di iodio elemento. Anche negli Stati Uniti, che pure non conoscono la diffusione del gozzo tiroideo nella misura che si riscontra in altri paesi, si verificano casi sempre piú frequenti di gozzo iatrogeno, cioè provocato da trattamento medico. Potenzialmente strumogeni sono infatti certi composti quali le solfonammidi prescritte per le infezioni al condotto urinario, i medicamenti a base di tiouracile normalmente somministrati per la cura dell'ipertiroidismo e diverse sostanze iodurate che vengono usate quali espettoranti nella terapia asmatica. Si sono avuti casi di morte prenatale di bambini

le cui madri avevano preso specialità di questo tipo: il preparato si diffonde nella placenta ed entra nel circolo fetale, provocando la morte *in utero* a causa dell'insorgere del gozzo nel feto ormai perfetto. Questi stessi medicamenti, concentrandosi nelle ghiandole mammarie, possono anche indurre il gozzo nei lattanti. Ma la conseguenza piú grave di certe terapie nella gravidanza è che le sostanze predette inibiscono il trasferimento della tiroxina materna al feto in formazione, mentre la tiroxina è assolutamente indispensabile allo sviluppo fisico e mentale dell'infante.

Diverse indagini epidemiologiche hanno consentito di accertare che in una data comunità il gozzo può insorgere spontaneamente, persistere per qualche tempo e quindi regredire, senza che sia intervenuta alcuna modificazione apparente nelle abitudini alimentari e di vita della popolazione. Per spiegare tali fenomeni epidemici è stata proposta l'eventualità di un'origine infettiva della disfunzione, ma la circostanza non è stata mai comprovata. Potrebbe darsi comunque che nella normale flora intestinale si manifesti una colonia di batteri iodochelanti, con conseguente impoverimento dello iodio alimentare a disposizione dell'organismo.

A proposito dell'intervento batterico, sia pure in modo sostanzialmente diverso, è stato possibile documentare almeno un caso di gozzo dovuto a microrganismi. Robert McCarrison, mentre studiava nel 1906 le popolazioni gozzute dell'Himalaya, nel corso delle frequenti visite a diversi villaggi della valle del fiume Gilgit, restò subito impressionato dal fatto che, mentre il paesino piú a monte denunciava un'incidenza minima di gozzi (12 %), la frequenza aumentava mano a mano che si scendeva a valle di villaggio in villaggio, finché nell'ultimo si riscontrava un'incidenza del 45 %. C'era poi un villaggio isolato i cui abitanti, pur vivendo come gli altri sulle rive del fiume, non bevevano acque fluviali: in questo villaggio il gozzo era sconosciuto. Allora McCarrison intraprese una sperimentazione controllata su trenta volontari suddivisi in due gruppi: il primo gruppo beveva l'acqua fangosa del fiume previa bollitura, mentre il secondo gruppo (comprendente lo stesso McCarrison) beveva la stessa acqua non bollita. Nello spazio di un mese quasi tutti i componenti del gruppo di McCarrison, al contrario di quelli dell'altro gruppo, cominciarono a manifestare l'insorgere del gozzo. Lo studioso ne concluse che la responsabilità dell'affezione doveva attribuirsi ai batteri. Le pessime condizioni igieniche della zona facevano sí che nel fiume andassero a finire tutti i rifiuti dei vari villaggi; la contaminazione delle acque aumentava quindi in relazione alla distanza dal villaggio piú a monte, con conseguente aumento della percentuale batterica nell'acqua consumata piú a valle. In seguito è stato dimostrato che certe colonie di *Escherichia coli*, batterio normalmente presente nel materiale fecale, sono in grado di produrre tiouracile.

Schema grafico dell'attività tiroidea. Lo ioduro catturato dalle cellule dei follicoli tiroidei viene trasformato in iodio dagli enzimi ossidanti delle cellule stesse. Lo iodio, combinandosi con la tirosina, forma gli ormoni tiroidei. Gli ormoni vengono immagazzinati in associazione con la tiroglobulina e, quando vengano richiesti, sono liberati dagli enzimi proteolitici e versati nel sangue. Il TSH stimola la produzione ormonica, favorendo con la sua azione la cattura dello ioduro e l'attività di tre serie di enzimi. Le sostanze strumogene interferiscono con la produzione ormonica. I tiocianati e i perclorati arrestano la cattura di ioduro (*1*), il tiouracile ostacola gli enzimi (*2*), le solfonammidi arrestano la combinazione con tirosina (*3*).

Il gozzo di grandi proporzioni si osserva spesso nelle regioni in cui il gozzo iodocarenziale è endemico. Si è avuta notizia di gozzi dal peso di un chilo e mezzo o due chili, talvolta penzolanti sul petto. L'illustrazione mostra un tipico burattino bergamasco, della metà del secolo scorso, «Giopin», e dimostra la diffusione del gozzo, in alcune regioni alpine, fino a risultare elemento di tipizzazione nella cultura popolare.

Nel caso del gozzo endemico si può parlare di un circolo vizioso. Le collettività povere a dieta alimentare invariata sono le piú esposte alla disfunzione e le piú vulnerabili alle sue conseguenze biologiche, sociali ed economiche. Laddove il gozzo da carenza di iodio sia endemico a una data comunità umana, è molto probabile che anche gli animali domestici siano affetti da ipotiroidismo. Gli ovini gozzuti spesso sono poveri di lana; nei bovini il gozzo provoca la sterilità, una scarsa produzione di latte, vitelli scarni e stenti; i cavalli gozzuti sono piú deboli; le galline sofferenti di scarsa attività tiroidea danno uova con il guscio povero di calcio, con conseguente rottura delle uova stesse e accresciuta mortalità del pollame. Le conseguenze dei mali che colpiscono il bestiame sono disastrose per la comunità povera la cui sopravvivenza dipenda quasi interamente dagli animali domestici.

L'effetto della topografia montagnosa e dell'isolamento sul cretinismo è evidentissimo. Ci sono infatti delle regioni in cui, pur essendo diffuso il gozzo, il cretinismo è quasi del tutto sconosciuto. È il caso, per esempio, della cosiddetta «fascia del gozzo» nella regione dei Grandi Laghi: è probabile che l'elevata mobilità demografica lungo questo tipico canale di migrazione verso ovest sia bastata a garantire quella varietà biologica e sociale che ha impedito al cretinismo di prendere piede.

Il cretinismo non è che l'esempio estremo delle conseguenze di una insufficiente disponibilità di tiroxina nei primi stadi dello sviluppo vitale. E poiché tutti i residenti di una data zona colpita da gozzo endemico sono potenzialmente esposti nella fase di sviluppo a un'erogazione subottimale dell'ormone, l'intera comunità può soffrire di effetti socialmente negativi e dannosi, essendo in gioco la motivazione psicologica, la spontaneità, la creatività, l'intelligenza degli individui, con conseguente ristagno sociale e ulteriore tendenza all'endogamia.

Prima o poi la scienza medica e le misure di sanità pubblica giungeranno a eliminare l'aspetto endemico del gozzo dovuto a carenza di iodio; c'è solo da sperare che quest'antichissima affezione benigna non venga soppiantata da una diversa disfunzione tiroidea, una disfunzione da era atomica, derivante dall'ingestione di rilevanti quantità di isotopi di iodio radioattivo dovuti al *fall-out* nucleare. Lo iodio si concentra nella tiroide, e qui la radioattività può danneggiare irreparabilmente le cellule interessate. Lo studio sul gozzo endemico dimostra la gravità di questo pericolo potenziale e gli effetti che ne potrebbero scaturire a danno dell'evoluzione dell'intera umanità.

Per ottenere i risultati sperimentali discussi in questo articolo è stato impiegato l'acceleratore lineare per elettroni dello Stanford Linear Accelerator Centre (SLAC), lungo due miglia. Il fascio di elettroni, muovendosi lungo una linea alimentata con klystron e guidato da magneti focalizzanti, raggiunge una energia massima di 21 miliardi di elettronvolt (GeV). Vicino all'estremità di tale linea il fascio di elettroni attraversa un « deflettore di fascio » prima di raggiungere le zone del bersaglio.

La struttura del protone e del neutrone

Il modo in cui i protoni e i neutroni diffondono gli elettroni di altissima energia fa avanzare l'ipotesi che le particelle « elementari » presentino una struttura interna come se fossero costituite da entità puntiformi

di Henry W. Kendall e Wolfgang K.H. Panofsky

Sessantacinque anni fa Ernest Rutherford osservò per la prima volta la diffusione di particelle alfa incidenti su un sottilissimo bersaglio metallico e dedusse che l'atomo non è un corpo omogeneo ma è costituito da elettroni negativi che circondano un nucleo centrale molto piccolo, pesante e carico positivamente. Da allora, i fisici di diversi laboratori hanno eseguito esperimenti di diffusione impiegando particelle di energia sempre maggiore, con l'intenzione di scoprire inizialmente la struttura atomica, poi la struttura nucleare e ora la struttura delle particelle che costituiscono il nucleo, cioè dei protoni e dei neutroni. Si vuole cioè scoprire se queste particelle « elementari » sono veramente corpi omogenei. I recenti esperimenti con elettroni di 21 miliardi di elettronvolt di energia, prodotti con l'acceleratore lineare di due miglia di lunghezza dello Stanford Linear Accelerator Center (SLAC) lasciano intravedere la possibilità che si ripeta quanto è avvenuto per l'atomo ma con un ordine di grandezza 100 000 volte minore.

I risultati ottenuti mostrano infatti che gli elettroni super veloci sono diffusi dai protoni e dai neutroni in un modo che nessuno aveva previsto. In prima ipotesi si potrebbe pensare che le particelle nucleari presentino una struttura interna complessa e siano costituite da entità puntiformi chiamate per ora « partoni ».

Secondo i risultati ottenuti i partoni presenterebbero alcune proprietà che precedentemente sono state attribuite a quelle particelle teoriche chiamate « quark ».

La conoscenza delle strutture interne del protone e del neutrone potrebbe rendere comprensibili le interazioni forti che garantiscono la stabilità dell'edificio nucleare e in definitiva dell'intero universo. L'interazione forte si manifesta nelle reazioni nucleari che tengono in vita le stelle e nelle reazioni che, su scala piú modesta, forniscono energia nucleare sia per scopi pacifici sia per scopi bellici. Sebbene questa interazione forte sia comunemente sfruttata nelle applicazioni tecnologiche, ancora oggi la sua natura e la sua origine non sono molto chiare.

I protoni e i neutroni, oltre a essere soggetti alla interazione forte, sono soggetti anche alla interazione elettromagnetica, che è circa 100 volte piú debole. Entrambe queste particelle nucleari si comportano come piccoli magneti ed entrambe posseggono una carica elettrica (anche se la carica totale del neutrone è nulla). Mentre l'interazione forte si manifesta solo quando le particelle interagenti sono molto vicine le une alle altre (a una distanza grosso modo pari al loro diametro: circa 10^{-13} centimetri), l'interazione elettromagnetica agisce anche per distanze infinite, diminuendo ovviamente di intensità proporzionalmente all'inverso del quadrato della distanza. Poiché il neutrone e il protone sono soggetti all'interazione elettromagnetica, essi diffondono gli elettroni inviati su di essi. L'andamento della diffusione che cosí si ottiene può dare una indicazione sulla loro struttura.

Poiché gli esperimenti che si effettuano a Stanford hanno caratteristiche del tutto analoghe a quelle degli esperimenti di Rutherford, è opportuno richiamare brevemente le tecniche da questi usate e i risultati ottenuti. Rutherford dispose un emettitore naturale di particelle alfa (particelle aventi una carica $+2$, identificate successivamente come nuclei di elio) in un contenitore – nel quale veniva praticato il vuoto – munito di un collimatore che permetteva di bombardare un bersaglio, costituito da un sottile foglio metallico, con un fascio di particelle ben definito. Nel contenitore era inoltre collocato un piccolo schermo di solfuro di zinco che scintillava quando veniva colpito da una particella alfa. Questo schermo poteva essere mosso in modo da intercettare le particelle diffuse a diversi angoli e le scintillazioni potevano essere contate a una a una con l'aiuto di un microscopio a basso ingrandimento. Ben presto, i due collaboratori di Rutherford, Hans Geiger e Ernest Marsden, notarono che le particelle alfa venivano diffuse a grandi angoli con una probabilità molto maggiore di quanto era stato previsto sulla base delle conoscenze correnti sulla struttura atomica. Si credeva infatti che negli atomi la carica elettrica fosse uniformemente distribuita, ma questa concezione non poteva interpretare gli ampi angoli di deflessione che, per essere spiegati, richiedevano invece l'esistenza di campi elettrici concentrati.

Rutherford concluse che « la carica positiva associata a un atomo è concentrata in una piccola zona centrale, detta nucleo, e che la carica negativa che la neutralizza è distribuita su una sfera di raggio avente dimensioni uguali a quelle dell'atomo ». Egli inoltre dedusse una legge matematica che descriveva la diffusione di una carica elettrica puntiforme, causata da una seconda carica elettrica pure puntiforme, supponendo che la forza agente tra le due particelle cariche fosse espressa dalla legge di Coulomb. Nota la carica e la massa delle particelle interagenti, Rutherford, combinando la legge di Coulomb con le leggi newtoniane del moto, mise in relazione la probabilità di diffusione secondo un dato angolo con l'energia della particella incidente. La probabilità di diffusione di un bersaglio costituito da un singolo atomo è chiamata « sezione d'urto di diffusione » ed è definita come l'area del fascio incidente nella quale l'influenza dell'atomo bersaglio provoca il processo osservato – in questo caso la diffusione.

La sezione d'urto non deve neces-

« Adone », l'acceleratore realizzato dai fisici italiani presso i Laboratori nazionali di Frascati del CNEN, è una macchina di concezione nuova e appartiene alla categoria dei cosiddetti « anelli di accumulazione ». Gli anelli di accumulazione per elettroni e positoni sono nati in Italia: il primo esemplare sperimentale (AdA) fu costruito presso gli stessi Laboratori di Frascati.
L'innovazione consiste nell'accelerare contemporaneamente due fasci di particelle di carica opposta (elettroni e positoni rispettivamente): i due fasci accelerati, catturati in una struttura magnetica circolare, percorrono la stessa orbita in verso opposto e si scontrano frontalmente in alcuni punti prestabiliti della macchina a una frequenza di circa 10^7 impatti al secondo.
Rispetto al metodo tradizionale di inviare una particella-proiettile contro un bersaglio fisso, lo studio della struttura delle particelle mediante l'urto di un proiettile e di un bersaglio che si muovono in verso opposto con la stessa quantità di moto presenta notevoli vantaggi. Infatti in questo caso il centro di massa del sistema proiettile-bersaglio risulta in quiete nel laboratorio: tutta l'energia posseduta dalle particelle è allora disponibile per l'interazione e il massimo impulso trasferito è esattamente uguale al doppio della quantità di moto posseduto da ciascuna particella prima dell'urto. Nell'urto di un proiettile con un bersaglio fisso, invece, parte dell'energia è associata semplicemente al moto del baricentro e non è quindi disponibile per l'interazione.
Nell'urto tra un elettrone accelerato e un protone fisso questo effetto potrebbe a prima vista sembrare trascurabile, visto che l'elettrone ha una massa circa duemila volte piú piccola di quella del protone: ma alle energie di cui stiamo parlando intervengono pesantemente effetti relativistici per cui per esempio gli elettroni accelerati a Stanford a 20 GeV risultano avere una massa quarantamila volte piú grande della loro massa a riposo. Come risultato, l'elettrone proiettile può trasferire al massimo un trecentesimo della sua energia a un elettrone bersaglio, e un terzo circa a un protone bersaglio.
« Adone », pur accelerando i fasci a una energia di soli 1,5 GeV (e quindi avendo dimensioni e costi proporzionatamente minori rispetto all'accelaratore di Stanford) permette di realizzare nell'urto elettrone-positone dei momenti trasferiti ben venti volte maggiori di quelli che nella stessa reazione si ottengono a Stanford, e solo due volte piú piccoli di quelli che a Stanford sono ottenibili nell'urto elettrone-protone.
Adone ha iniziato a essere utilizzato per la sperimentazione intorno alla fine del 1969 e, a partire dal settembre 1970 i risultati dei primi esperimenti sono stati comunicati ai congressi internazionali. Contrariamente a quanto si può credere, essi non riguardano soltanto la struttura dell'elettrone e del positone. Nell'impatto fra elettrone e positone (l'uno antiparticella dell'altro) può infatti avere luogo la conversione completa in massa di tutta l'energia che le particelle hanno accumulato nel processo di accelerazione, il che si manifesta attraverso la creazione di nuove particelle (due o piú ioni, coppie protone-antiprotone, etc.).
Questi processi hanno luogo principalmente attraverso la formazione di un fotone virtuale come stato intermedio, e risultano particolarmente adatti allo studio della struttura temporale – anziché spaziale, come negli esperimenti di Stanford – sia del fotone stesso sia delle particelle prodotte. In questo senso, gli esperimenti di Frascati sono complementari a quelli di Stanford.
Le sezioni d'urto misurate a Frascati sono, come quelle misurate a Stanford, di almeno un ordine di grandezza piú grandi di quelle previste dai modelli teorici precedentemente esistenti. Ciò indica da un lato che il fotone, che si sapeva vivere una parte della sua vita nella forma virtuale di mesone vettoriale (ρ, ω, Φ), nella sua evoluzione temporale passa attraverso stati virtuali di massa piú elevata, che appaiono essere, anziché particelle, stati continui di materia.
D'altro canto il valore delle sezioni d'urto misurate appare sorprendente in accordo con la sezione d'urto prevista per produzione di una coppia particella-antiparticella, ove queste siano prive di struttura. Cosí come la piú suggestiva interpretazione degli esperimenti di Stanford è che il protone sia formato di partoni, suoi costituenti privi di struttura, l'interpretazione parallela degli esperimenti di Frascati è che nell'interazione elettrone-positone vengano prodotte coppie di partone-antipartone, i quali interagendo fra loro, una volta creati, producono le particelle effettivamente osservate dagli apparati di rivelazione.
L'esistenza dei partoni, è per ora solo una ipotesi, e per confermarla è necessario ancora molto lavoro di indagine sperimentale. Ma è diffusa fra i fisici la sensazione che ci si trovi davanti a una porta che sta per essere aperta.

V. S.

sariamente essere messa in relazione alla « vera » dimensione fisica della particella bersaglio, ma rappresenta piuttosto una misura della forza esercitata sulla particella incidente da parte della particella bersaglio.

La sezione d'urto viene determinata sperimentalmente per i diversi angoli (misurati relativamente all'asse del fascio incidente) e i risultati possono essere confrontati con le previsioni teoriche. La formula di Rutherford valuta la sezione d'urto di diffusione, note che siano la massa m e la carica della particella incidente, la massa e la carica della particella bersaglio, la velocità v della particella incidente e l'angolo di diffusione ϑ. La formula dipende direttamente dalla particolare combinazione di queste variabili, combinazione che descrive il vettore differenza, q, fra la quantità di moto iniziale e finale della particella diffusa $q = 2mv$ (sin $\vartheta/2$). La quantità q viene anche chiamata « quantità di moto trasferita » (*si veda la figura nella pagina a fronte*). Nella formula si considera che le particelle interagenti siano punti matematici, senza dimensioni e senza forma. In generale, tuttavia, la sezione d'urto di diffusione dipenderà oltre che dalle caratteristiche delle forze (per esempio da come la loro intensità dipende dalla distanza) e dalle leggi del moto delle particelle (che possono non essere newtoniane o relativistiche), anche dalla struttura interna caratteristica.

Nei processi di diffusione descritti dalla meccanica quantistica la quantità di moto trasferita è una quantità molto importante, in quanto determina la dimensione della grandezza in studio. Secondo la meccanica quantistica, se una particella possiede una data quantità di moto p a essa è pure associata una data lunghezza d'onda λ. La relazione che lega queste due quantità è la seguente, $\lambda = h/p$, dove h è una quantità molto piccola ($6{,}6 \times 10^{-27}$ erg secondo) nota come costante di Planck. La precisione con la quale può essere localizzata una particella è limitata dall'onda associata; la probabilità di trovare la particella in un dato punto dipende dalle caratteristiche del « pacchetto d'onda » che descrive il moto della particella. L'avvicinamento di due particelle dà luogo a una interazione (cioè, da un punto di vista sperimentale si manifesta la diffusione di una particella da parte dell'altra), e questa interazione provoca un trasferimento della quantità di moto tra le due particelle. È quindi logico che l'indeterminazione Δx con la quale è possibile esaminare una struttura viene a dipendere da q, quantità di moto trasferita nella collisione, secondo la relazione $\Delta x = h/q$. Da questa relazione si deduce che, per evidenziare i piccoli dettagli eventualmente presenti nella particella, è necessario che la quantità q sia piú grande possibile in modo che la lunghezza d'onda λ sia piú piccola possibile.

Negli esperimenti di diffusione che vengono eseguiti a Stanford la quantità di moto è misurata in GeV/c, dove con il simbolo « GeV » si indicano i gigaelettronvolt (10^9 elettronvolt, cioè un miliardo di elettronvolt) mentre c rappresenta la velocità della luce. Un elettrone di 20 GeV si muove con una velocità che differisce dalla velocità della luce solo per uno su tre miliardi. In queste condizioni l'energia della particella, espressa in GeV, e la sua quantità di moto associata, espressa in GeV/c, sono equivalenti.

ELETTRONE DIFFUSO
NUCLEONE BERSAGLIO
ELETTRONE INCIDENTE
p' q p ϑ

La quantità di moto trasferita è una grandezza fondamentale nella trattazione teorica dei processi di diffusione, ed è definita come il vettore differenza (q) fra la quantità di moto iniziale (p) e la quantità di moto finale (p') della particella diffusa. La relazione che esprime questa probabilità è $q = 2mv$ (sin $\vartheta/2$), dove m rappresenta la massa della particella incidente, v la sua velocità e ϑ l'angolo di diffusione. Nella diffusione elastica il nucleone bersaglio è soggetto a rinculo; nella diffusione anelastica può o disintegrarsi, originando altre particelle, o rimanere in uno stato eccitato.

I due tipi di diffusione

La diffusione degli elettroni può essere « elastica » o « anelastica ». Nella diffusione elastica la particella bersaglio rincula come se fosse una palla da biliardo, restando tuttavia nello stesso stato di energia interna nel quale si trovava prima della collisione. Nella diffusione anelastica la particella bersaglio può disintegrarsi oppure passare in uno stato eccitato, cioè in uno stato diverso da quello iniziale. I due processi non possono coesistere: uno esclude l'altro. Entrambi forniscono interessanti informazioni sulla struttura della particella bersaglio. Incominceremo con il parlare della diffusione elastica.

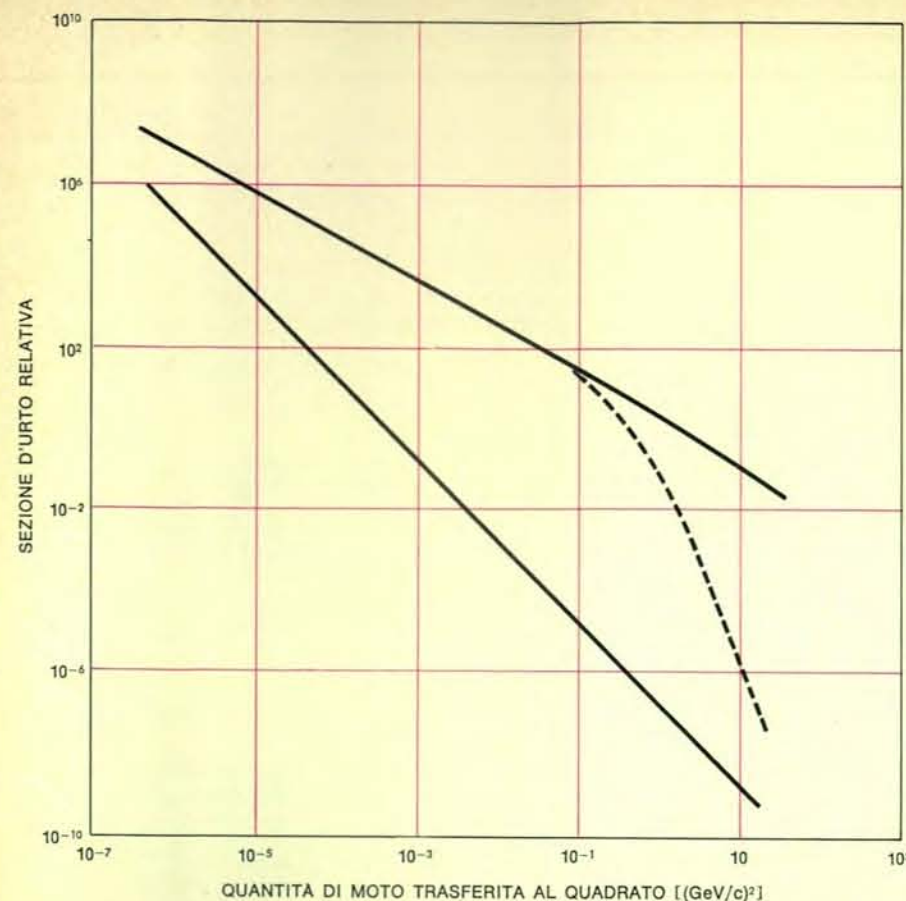
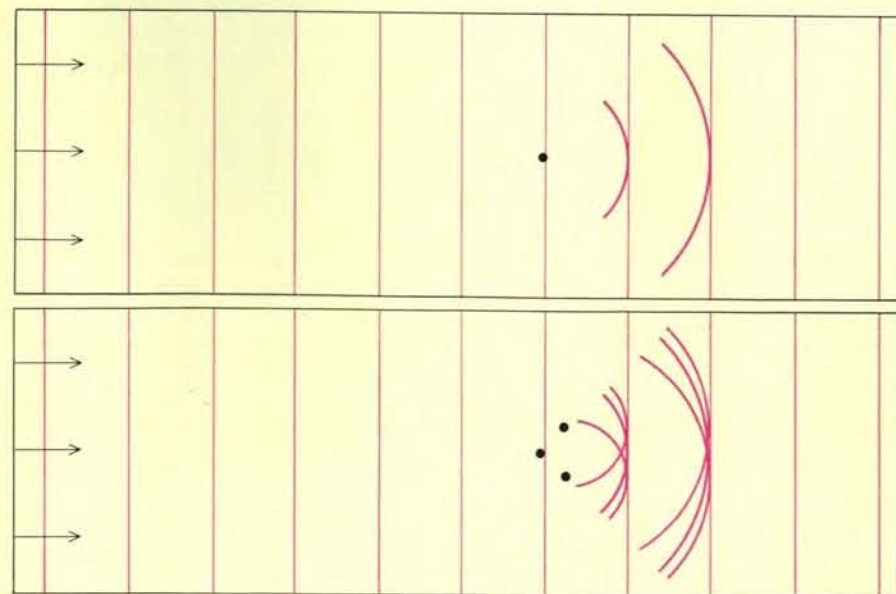
La relazione di Rutherford non è in grado di descrivere adeguatamente la diffusione elastica degli elettroni di energia elevata per due fondamentali motivi. In primo luogo le velocità in gioco sono cosí elevate che è necessario impiegare la teoria quantistico-relativistica per descrivere la natura ondulatoria e il comportamento della particella incidente e di quella bersaglio. In secondo luogo, gli elettroni presentano uno « spin », cioè un momento angolare, come se fossero soggetti a un moto rotatorio su se stessi. Una formula piú adeguata che può essere impiegata in questo caso è conosciuta come sezione di urto di Mott. Fatta eccezione per il termine che tiene conto dello spin dell'elettrone, l'equazione di Mott si riduce all'equazione di Rutherford in tutti quei casi in cui la velocità della particella incidente è molto inferiore alla velocità della luce, come d'altronde si verificava nelle esperienze di Rutherford (*si veda la figura a pag. 72 in alto*). Poiché Rutherford non sapeva che i suoi esperimenti dovevano essere descritti dalla meccanica quantistica, è solo un caso veramente fortunato che la sua relazione descriva correttamente la diffusione a basse energie.

La stessa equazione di Mott deve essere modificata se l'elettrone viene diffuso invece che da una carica puntiforme, da un oggetto di dimensioni finite (*si veda la figura in basso a pag. 72*). In questo caso ogni segmento del fronte d'onda che rappresenta l'elettrone viene diffuso separatamente da ciascuna sotto-unità di carica contenuta nella particella bersaglio. Le singole onde diffuse dalle sotto-unità si ricombinano e danno origine all'onda diffratta che rappresenta l'elettrone diffuso. Alcune parti delle onde si sommano costruttivamente, altre invece interferiscono, annullandosi le une con le altre. La sezione d'urto di diffusione elastica, per una particella carica di dimensioni finite, è generalmente inferiore alla sezione d'urto per una carica puntiforme. Il rapporto fra la sezione d'urto di diffusione dovuta a una particella di dimensioni finite rispetto a quella dovuta a una particella puntiforme è uguale al quadrato di una grandezza chiamata fattore di forma, e indicata con la lettera F.

La relazione che determina il fattore di forma è ottenuta valutando il percorso residuo, nella particella bersaglio, delle onde parziali dopo che queste sono state diffuse dalle sotto-unità di carica. La relazione dipende essenzialmente dalla quantità di moto trasferita, q, che rappresenta il vettore differenza fra le quantità di moto dell'elettrone incidente e di quello emergente. Per valori sufficientemente elevati di q, il fattore di forma risulta piú sensibile ai dettagli della struttura del bersaglio; se invece q è piccolo, l'esperimento fornisce pochi particolari.

Sarebbe interessante studiare la struttura interna di un nucleone (protone o neutrone) a distanze inferiori al suo raggio, che è di circa 0,8 fermi (un fermi è uguale a 10^{-13} cm). Per avere una risoluzione di circa 0,1 fermi, è necessario che la quantità di moto trasferita sia di 2 GeV/c. Negli attuali esperimenti, il limite pratico è di circa 5 GeV/c, ed è perciò sufficientemente piccolo per fornire sostanziali informazioni sulla struttura del protone. Se sarà possibile conoscere il fattore di forma per diversi valori della quantità di moto trasferita, sarà possibile ricostruire la distribuzione di carica della particella bersaglio.

La determinazione della distribuzione di carica su una particella, per esempio il protone, per mezzo della diffusione di elettroni è molto simile alla ricostruzione della struttura di un cristallo eseguita attraverso una complicata figura di diffrazione prodotta in seguito al bombardamento con raggi X. Inoltre il problema della diffusione di elettroni è molto piú complesso, particolarmente quando la velocità del protone di rinculo è prossima alla velocità della luce. Gli effetti della relatività sul moto del protone introducono delle ambiguità che complicano l'analisi della distribuzione spaziale della carica.

Una ulteriore complicazione viene introdotta dallo spin del protone, che genera un momento magnetico. Di conseguenza l'elettrone incidente può interagire anche con il campo magnetico del protone oltre che con il suo campo elettrico. Poiché il campo magnetico può anche presentare una distribuzione spaziale finita, si deve introdurre un secondo fattore di forma, F_m, diverso dal fattore di forma elettrico F_e. In definitiva la relazione originale di Rutherford

La sezione d'urto è definita come l'area del fascio incidente nella quale l'influenza dell'atomo bersaglio dà origine a un certo tipo di interazione. In questa figura viene rappresentata la sezione d'urto di diffusione di un elettrone da parte di un nucleone bersaglio valutata con la relazione di Rutherford (*linea inferiore*) e con la relazione di Mott (*linea superiore*). La relazione di Mott coincide, alle basse energie dell'elettrone incidente, con quella di Rutherford se si esclude il termine che tiene conto dello spin dell'elettrone. La curva tratteggiata rappresenta la relazione di Mott per la diffusione di un protone di dimensioni finite. Le curve corrispondono a un angolo di 20°.

Se si suppone che l'elettrone venga diffuso non da una particella carica puntiforme (*sopra*), ma da una particella di dimensioni finite (*sotto*), qui rappresentata come costituita da tre punti, è necessario modificare la formula che descrive la diffusione. Nel caso che la particella bersaglio abbia dimensioni finite, ogni segmento del fronte d'onda che rappresenta l'elettrone viene diffratto separatamente da ciascuna sotto-unità di carica. Le onde parziali diffuse da queste sotto-unità di carica elettrica si ricombinano tra loro dando così origine all'onda emergente che rappresenta l'elettrone diffuso.

deve essere modificata per tener conto di questi fatti: la particella incidente e la particella bersaglio presentano uno spin, la particella bersaglio ha dimensioni finite, la velocità di collisione è tanto elevata da introdurre effetti relativistici, e infine il moto di entrambe le particelle deve essere descritto con la meccanica ondulatoria e non con la meccanica classica (*si veda la figura nella pagina a fronte*).

Questa trattazione alquanto elaborata non deve spaventare perché il processo di diffusione elettronica è fondamentalmente un processo semplice. Esso infatti, per mezzo delle ben note forze elettromagnetiche, permette di esplorare la struttura, tuttora sconosciuta, delle particelle subnucleari; questo in contrasto con quegli esperimenti (interessanti per altre ragioni) nei quali si fanno collidere tra loro particelle di struttura sconosciuta, come per esempio gli esperimenti di diffusione protone-protone o pione-pione. Per quanto si sa, gli elettroni si comportano come punti carichi e interagiscono, negli esperimenti di diffusione, solo con forze di tipo elettromagnetico. (È naturalmente vero che gli elettroni interagiscono anche per mezzo di interazioni « deboli », forze queste che giocano un ruolo importante nei processi di decadimento radioattivo, ma poiché le interazioni deboli sono circa 10^{10} volte meno intense delle interazioni elettromagnetiche, esse possono essere trascurate negli sperimenti di diffusione elettronica).

Le leggi dell'elettricità e del magnetismo, espresse dalle equazioni della elettrodinamica quantistica, costituiscono, in fisica, il solo e unico esempio nel quale una unica descrizione quantitativa è risultata valida per l'intera gamma di esperimenti eseguiti per verificarla, dalle dimensioni cosmiche fino a 10^{-15} cm. È quindi ben giustificato l'interesse per questo particolare tipo di forze.

L'acceleratore lungo due miglia

Prima di discutere i risultati ottenuti con gli esperimenti di diffusione elastica e anelastica, eseguiti per mezzo degli elettroni accelerati a Stanford, è opportuno descrivere brevemente questo acceleratore e le tecniche impiegate. Un tubo a vuoto, lungo 2 miglia (circa 3 km), alimentato da 245 klystron che forniscono l'energia elettromagnetica, permette di accelerare gli elettroni del fascio fino a un'energia massima di 21 GeV. Lungo questo percorso il fascio viene focalizzato da « lenti » magnetiche, distanti le une dalle altre 100 metri. Alla fine del suo percorso il fascio passa attraverso un « ripulitore » finale, costituito da magneti e da diaframmi che definiscono esattamente le sue dimensioni e la sua energia, prima che esso colpisca il bersaglio. Questi esperimenti di diffusione richiedono bersagli costituiti da idrogeno o deuterio e un sistema per selezionare e identificare gli elettroni diffusi ai diversi angoli e per misurare la loro quantità di moto in presenza di molte altre particelle, prodotte in seguito alla collisione degli elettroni con i nuclei del bersaglio.

I protoni bersaglio sono ottenuti da idrogeno liquido contenuto in un particolare recipiente. Il nucleo del normale idrogeno è costituito da un unico protone. Impiegando deuterio (idrogeno pesante) liquido si può disporre di bersagli di neutroni liberi. Il nucleo del deuterio è infatti costituito da un neutrone e da un protone e, con buona approssimazione, la diffusione prodotta dai nuclei di deuterio è data dalla somma della diffusione prodotta singolarmente dai neutroni e dai protoni. A causa della elevata intensità dei fasci impiegati, i bersagli, costituiti da gas liquefatti, devono essere raffreddati con continuità, per esempio per mezzo di uno scambiatore di calore, e questo non solo per prevenire un processo di ebollizione ma anche per ridurre al minimo le variazioni di densità che falserebbero completamente i risultati.

Per separare e classificare gli elettroni emergenti dal bersaglio sono disponibili, presso il laboratorio di Stanford, tre spettrometri magnetici che focalizzano gli elettroni in un sistema di rivelatori. Questi spettrometri devono presentare un'altissima risoluzione sia energetica che angolare, dato che devono consentire di distinguere gli elettroni diffusi elasticamente da quelli diffusi anelasticamente e risolvere la complicata struttura degli spettri energetici degli elettroni diffusi anelasticamente.

Nel processo di diffusione anelastica, in seguito alla collisione, possono essere prodotti uno o più pioni. Poiché la energia necessaria per creare un pione è di 139 MeV (milioni di elettronvolt), la risoluzione richiesta deve essere notevolmente migliore del rapporto fra 139 MeV e l'energia della particella incidente che, abbiamo visto, può superare i 20 GeV. È quindi necessaria una risoluzione energetica dello 0,7 per cento. Un'analisi analoga eseguita per la cinematica della collisione richiede che la risoluzione angolare sia una frazione di milliradiante, cioè una frazione di tre primi di grado. Tuttavia quello che più conta è la precisione sull'angolo relativo e sull'energia relativa dell'elettrone incidente e diffuso; pertanto è necessario che siano garantite, anche per il fascio incidente, le risoluzioni angolari ed energetiche richieste per gli spettrometri che analizzano il fascio diffuso.

Gli spettrometri sono grosse e complicate apparecchiature costituite da lenti magnetiche e da magneti di deflessione che deviano verticalmente gli elettroni diffusi e successivamente li focalizzano in un punto. La deflessione verticale operata fornisce un'indicazione della quantità di moto dell'elettrone; la posizione orizzontale fornisce un'indicazione dell'angolo di diffusione. Centinaia di contatori, equivalenti allo schermo scintillatore di solfuro di zinco impiegato da Rutherford nei suoi esperimenti, identificano l'angolo e la quantità di moto di ciascun elettrone. I contatori sono costituiti da strette barre di materiale plastico trasparente che presenta la caratteristica di scintillare per tempi brevissimi quando viene colpito da una particella di elevata energia. Ogni barra è collegata a un tubo fotomoltiplicatore che trasforma il segnale luminoso in un segnale elettrico.

I segnali provenienti dai contatori e da altri sistemi di identificazione di particelle, vengono elaborati e successivamente analizzati da un grosso calcolatore. Il calcolatore opera in linea, immagazzina i dati per le successive analisi più dettagliate e contemporaneamente fornisce semplici analisi parziali. Oltre a questi risultati il calcolatore fornisce anche informazioni sullo stato di funzionamento delle apparecchiature ed esegue molte operazioni di controllo, quali per esempio la regolazione della corrente per i magneti degli spettrometri, il controllo dell'intensità del fascio e di altri parametri particolarmente importanti.

Fattori di forma del nucleone

Gli esperimenti di diffusione elastica eseguiti dai fisici del Caltech, del MIT e dello SLAC, hanno permesso la determinazione dei quattro fattori di forma che descrivono la struttura del protone e del neutrone, anche se la qualità e la quantità dei dati finora raccolti risultano alquanto fluttuanti. Le misure più accurate sono quelle che forniscono il valore del fattore di forma magnetico del protone (*si veda la figura a pag. 75*). Il fattore di forma magnetico del neutrone, ottenuto sottraendo ai valori relativi alla diffusione prodotta dal deuterio quelli relativi alla diffusione attribuita al protone, presenta, entro gli errori che in questo caso sono maggiori, un andamento analogo a quello del protone stesso. Il fattore di forma elettrico per il protone assomiglia al fattore di forma magnetico; si deve tenere conto che in questo caso i risultati sono stati determinati per un intervallo delle variabili molto minore. Si è osservato che il fattore di forma elettrico per il neutrone è praticamente nullo, tuttavia gli errori delle misure eseguite sono molto grandi.

Ci si potrebbe chiedere perché gli elettroni vengano diffusi dal neutrone, anche se il neutrone non possiede carica elettrica. La risposta può essere divisa in due parti. Innanzi tutto lo spin del neutrone genera un momento magnetico; e questo da solo spiegherebbe

SEZIONE D'URTO DI RUTHERFORD $\sigma_R = \frac{(2e^2 m)^2}{q^4}$	e = CARICA DELL'ELETTRONE m = MASSA DELL'ELETTRONE $q = 2\sqrt{pp'}\ \mathrm{sen}\ \vartheta/2$ p = QUANTITÀ DI MOTO INIZIALE DELL'ELETTRONE p' = QUANTITÀ DI MOTO FINALE DELL'ELETTRONE
SEZIONE D'URTO DI MOTT $\sigma_M = \frac{(2e^2\ E'/c^2)^2}{q^4} \cdot \frac{E'}{E}\left(\cos^2 \frac{\vartheta}{2}\right)$	ϑ = ANGOLO DI DIFFUSIONE DELL'ELETTRONE E = ENERGIA INIZIALE DELL'ELETTRONE E' = ENERGIA FINALE DELL'ELETTRONE c = VELOCITÀ DELLA LUCE
SEZIONE D'URTO DI ROSENBLUTH $\sigma = \sigma_M \left(\frac{F_e^2 + \tau F_m^2}{1+\tau} + 2\tau F_m^2 \tan^2 \frac{\vartheta}{2}\right)$	F_e = FATTORE DI FORMA ELETTRICO F_m = FATTORE DI FORMA MAGNETICO $\tau = q^2/4 M^2 c^2$ M = MASSA DEL NUCLEONE

La sezione d'urto di diffusione viene descritta con crescente complessità, passando dalla relazione di Rutherford alla relazione di Mott e infine alla relazione di Rosenbluth. Quest'ultima relazione tiene conto delle seguenti condizioni: esistenza dello spin sia per la particella incidente sia per la particella bersaglio, dimensioni finite per la particella bersaglio, introduzione di effetti relativistici a causa delle elevate velocità di collisione, descrizione del moto delle particelle mediante la meccanica ondulatoria.

la diffusione descritta dal fattore di forma magnetico. In secondo luogo, quella corrente elettrica che è l'artefice del magnetismo del neutrone è in grado di produrre, nella particella stessa, accumuli localizzati di carica anche se la carica totale della particella è nulla. Tali accumuli danno origine a una diffusione elettrica ogni qualvolta il valore della quantità di moto trasferita è maggiore di zero. Di conseguenza la diffusione elastica di elettroni non solo risente della carica globale e del momento magnetico del neutrone, ma mette anche in evidenza i fenomeni che avvengono nel suo interno.

Gli esperimenti indicano che la struttura magnetica del neutrone e del protone sono quasi identiche e che il valore numerico della diffusione è proporzionale alle proprietà magnetiche di ciascuna particella, determinate per mezzo di esperimenti statici. In altre parole, le curve dei fattori di forma magnetici relative alle due particelle risultano, dai dati sperimentali, identiche nella forma. Molto probabilmente è significativo il fatto che la diffusione elettrica dovuta al protone sia proporzionale alla diffusione magnetica, in tutto l'intervallo analizzabile sperimentalmente. Questo fatto suggerisce che la distribuzione della carica elettrica, nel protone, deve essere messa direttamente in relazione alla struttura magnetica.

Le onde parziali, diffuse, creano una figura di diffrazione simile alla figura dell'ombra che si forma quando raggi di luce paralleli colpiscono il bordo di un oggetto. Se il bordo dell'oggetto è netto, la figura è costituita da righe alternativamente chiare e scure. Allo stesso modo, se il protone fosse un corpo con una superficie ben definita, si dovrebbe notare, nella curva del fattore di forma, un andamento meno continuo e regolare di quello che in realtà si osserva. Evidentemente quindi il protone presenta una superficie increspata irregolarmente. I dettagli della curva forniscono il raggio medio del protone, che risulta di circa 0,8 fermi.

Particelle reali e particelle virtuali

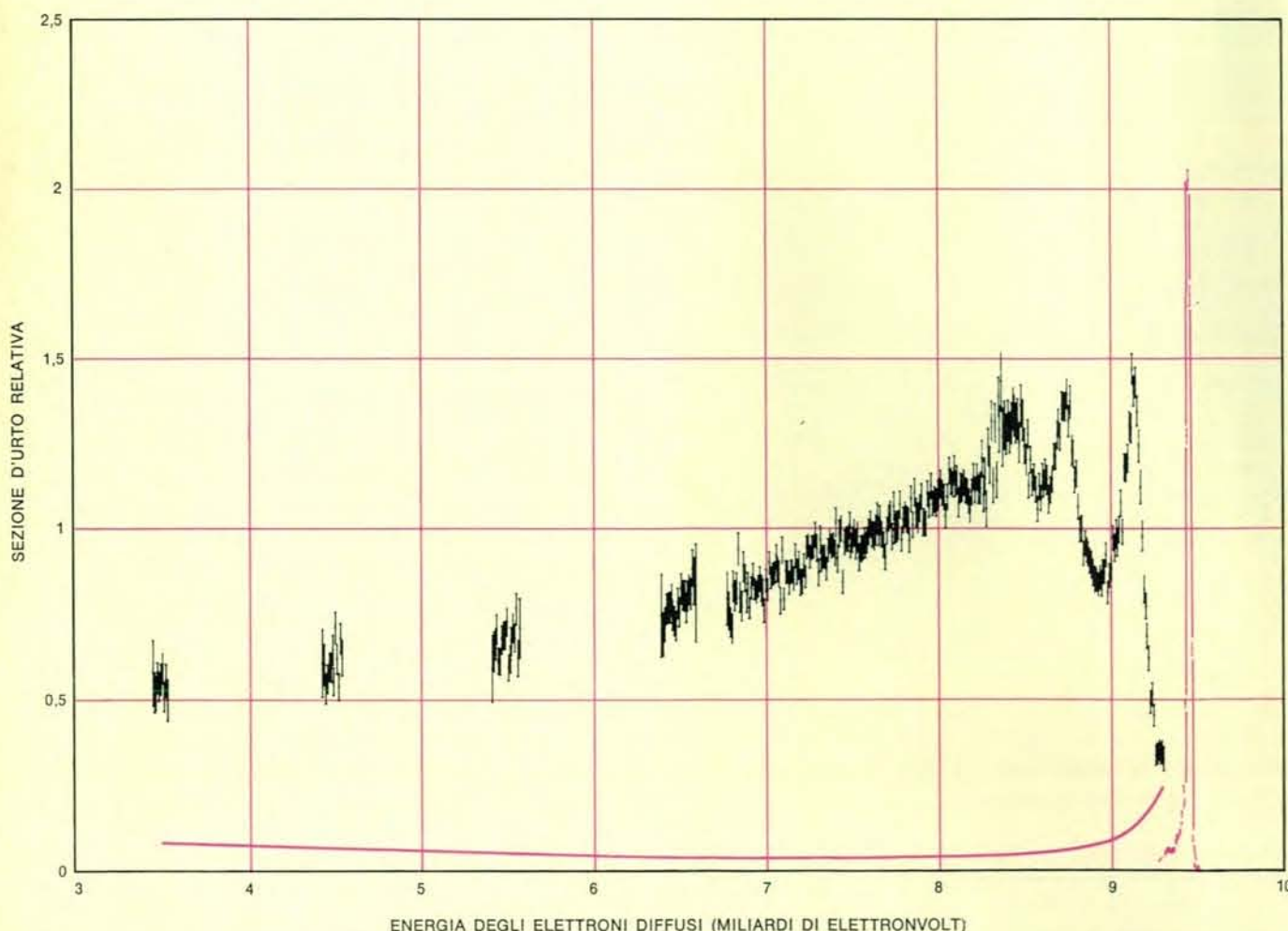
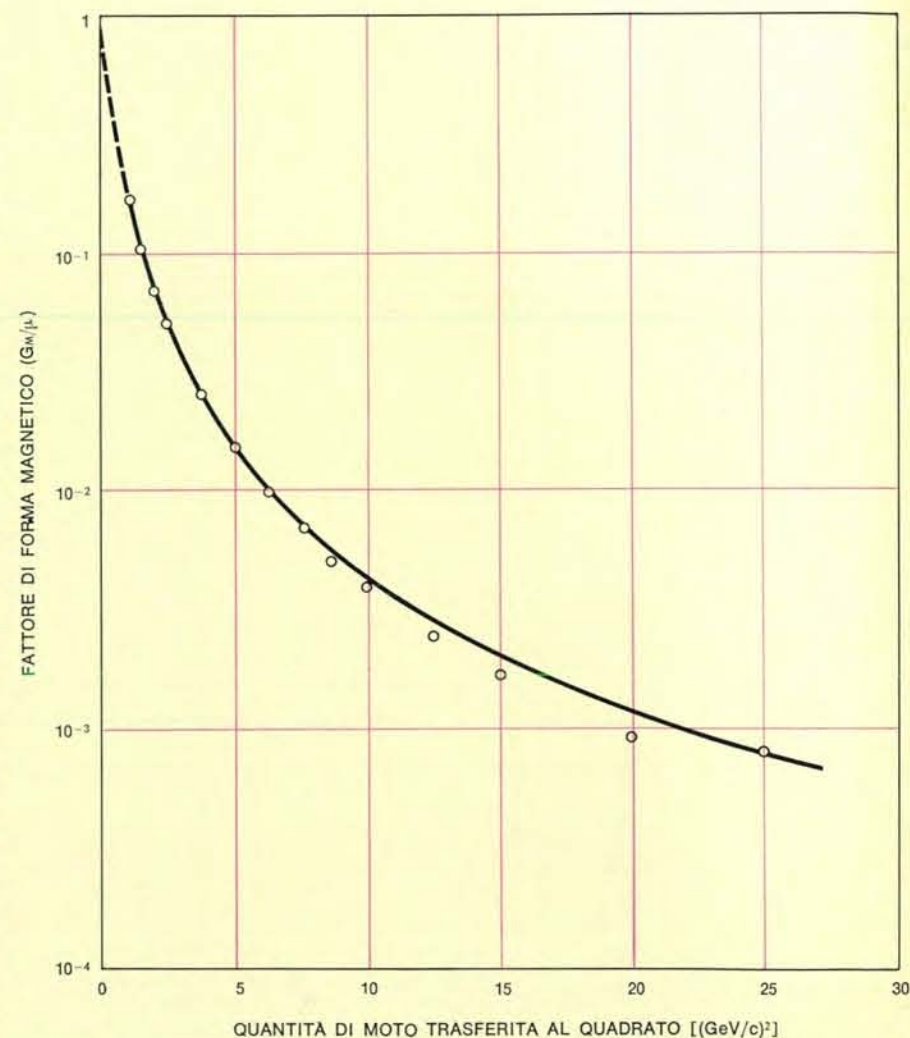
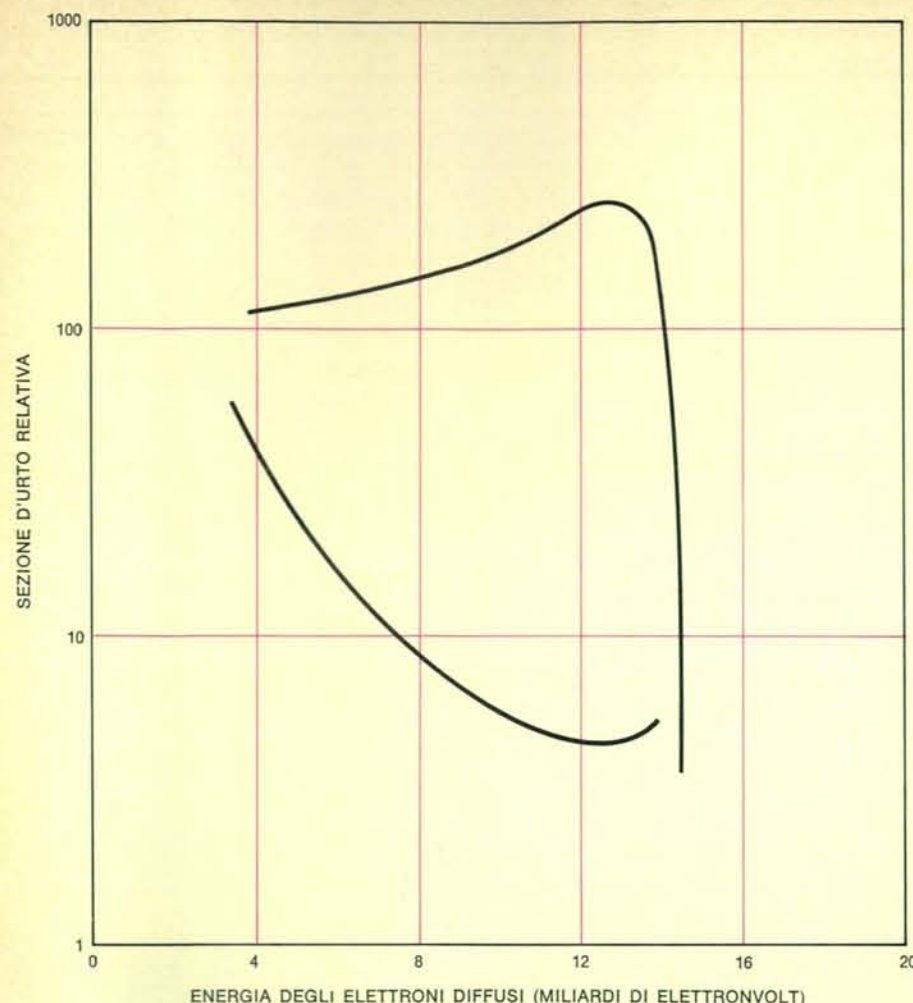
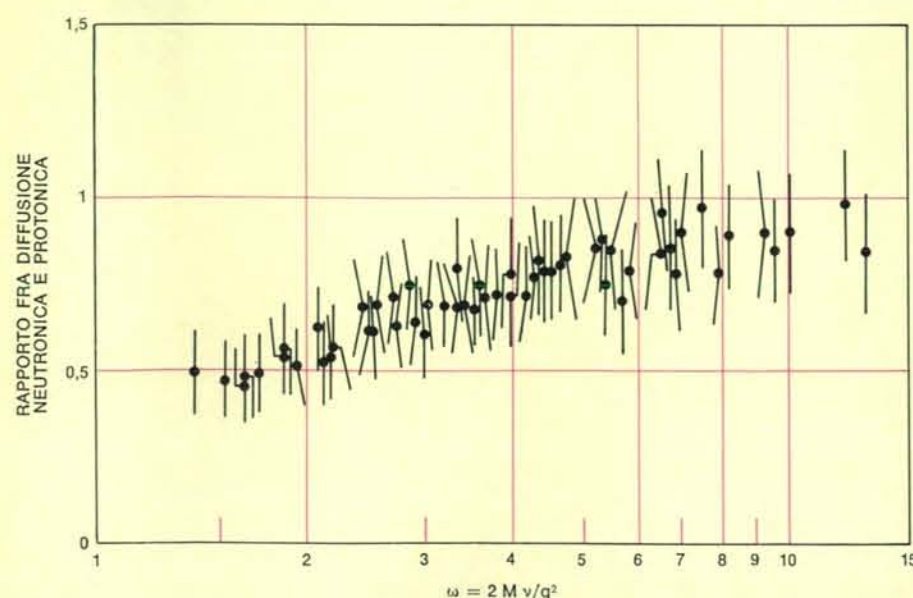
Per i fisici è stata una notevole sorpresa l'aver scoperto che la curva, riportata nella pagina a fronte, che rappresenta il fattore di forma magnetico del protone è notevolmente regolare entro un grandissimo intervallo delle variabili sperimentali. La sezione di urto sperimentale di diffusione, che dipende dal quadrato del fattore di forma moltiplicato per la formula di Mott, relativa a diffusione prodotta da bersagli puntiformi, decresce di un fattore 10^{12} nell'intervallo nel quale sono state eseguite le misure. Le sezioni d'urto relative ai punti piú bassi della curva in questione, risultano estremamente piccole; la piú piccola sezione d'urto misurata è di circa 2×10^{-39} centimetri quadrati per steradiante, e questo significa che nelle condizioni sperimentali, in un rivelatore è stato diffuso un solo elettrone su 10^{18} elettroni incidenti sul bersaglio. La diffusione diminuisce proporzionalmente alla quarta potenza della quantità di moto trasferita. Questa diminuzione cosí rapida costituisce un enigma per la fisica delle alte energie. Per comprendere il perché questo sia un enigma e come possa essere spiegato è necessario aprire una breve parentesi sul concetto di « particella virtuale ».

Il concetto di particella virtuale deve essere considerato in relazione al principio di indeterminazione enunciato da Werner Heisenberg piú di 40 anni fa. Nella descrizione ondulatoria della materia è impossibile determinare simultaneamente e in modo assolutamente preciso la lunghezza d'onda associata a una particella e la sua quantità di moto. Il principio di Heisenberg mette in relazione l'indeterminazione della misura della lunghezza d'onda della particella, Δx, con l'indeterminazione della quantità di moto della stessa particella, Δp. Il prodotto delle due indeterminazioni risulta proporzionale alla costante di Planck h ($\Delta p \cdot \Delta x \simeq h$). Parallelamente è possibile scrivere una relazione di indeterminazione fra la misura della energia della particella, ΔE, e l'indeterminazione temporale, Δt, della misura stessa. Tale relazione risulta essere $\Delta E \cdot \Delta t \simeq h$.

Secondo la teoria della relatività la massa e l'energia sono equivalenti, e questa equivalenza è espressa dalla relazione di Einstein $E = mc^2$. Si può perciò supporre che in un intervallo di tempo Δt molto breve, qualsiasi quantità di energia ΔE possa essere trasformata in una massa m, equivalente alla massa a riposo di qualche particella, a condizione, naturalmente, che il prodotto di ΔE per Δt non superi h. In altre parole, in un sistema, possono comparire ed esistere per periodi di tempo brevissimi e non misurabili, una o piú particelle senza che sia violato il principio di indeterminazione. In un certo senso l'esistenza di queste particelle è « nascosta » da una non eliminabile indeterminazione della nostra conoscenza del sistema. Le particelle che compaiono in questo modo sono chiamate particelle virtuali; esse non possono essere osservate direttamente come si possono invece osservare le particelle reali.

Diversi modelli che descrivono l'interazione fra l'elettrone e il protone introducono il fotone (quanto di energia) quale portatore di forze elettromagnetiche. Anche il fotone può essere virtuale o reale. I fotoni reali costituiscono i pacchetti d'onda che trasportano energia da una sorgente radiante (quale per esempio una stella) a un assorbitore (per esempio al pigmento visivo dell'occhio). Nell'elettrodinamica quantistica le forze elettromagnetiche che agiscono fra due (o piú) cariche in moto sono attribuite all'emissione e all'assorbimento di fotoni virtuali. Pertanto nella diffusione elettronica un fotone virtuale, emesso dall'elettrone, interagisce ed è assorbito dalla carica elettrica e dal campo magnetico del protone. I fotoni virtuali possono trasportare energia e quantità di moto in qualsiasi proporzione, a differenza di quanto avviene per i fotoni reali per i quali tali quantità devono essere univocamente correlate.

Sebbene possa sembrare che le particelle virtuali violino le leggi fondamentali di conservazione, la violazione è strettamente limitata a quei campi nei quali si applica il principio di indeterminazione. Per esempio questo principio non si applica alla conservazione della carica. Di conseguenza non è possibile che un singolo elettrone virtuale compaia da solo nel vuoto; esso deve essere sempre accompagnato da una particella di segno opposto, cioè da un positone.

Esiste una categoria di particelle instabili, i mesoni neutri, i cui componenti sono, sotto molti aspetti, simili ai fotoni, con due importanti eccezioni: i mesoni posseggono una massa e generano interazioni forti. In questa categoria si distingue il mesone ρ, che possiede una massa equivalente a circa 750 MeV (la massa del protone è equivalente a 939 MeV). I mesoni ρ possono essere prodotti in laboratorio come particelle reali, ed è possibile rivelare i loro prodotti di decadimento. I mesoni neutri possono anche essere prodotti, come

Questo spettro di diffusione che comprende sia gli eventi elastici (*in colore*), sia gli eventi anelastici (*in nero*), è stato ottenuto in seguito al bombardamento di protoni da parte di un fascio di elettroni aventi una energia di 10 GeV. Il picco elastico (*a destra*) è ridotto per motivi grafici di un fattore cinque; l'asimmetria di forma è dovuta alla possibile emissione di raggi X « molli » da parte degli elettroni. Questi raggi X « rubano » una certa parte di energia e deformano, verso le basse energie, il picco elastico. I picchi piú piccoli o picchi secondari presenti nello spettro anelastico corrispondono agli stati eccitati del protone, cioè alle risonanze. Alla sinistra di questi picchi secondari si nota uno spettro continuo e regolare chiamato « continuo ». Aumentando l'energia delle particelle incidenti le risonanze tendono a sparire mentre il continuo rimane.

I ricercatori del Caltech, del MIT e dello SLAC, hanno notato che il fattore di forma magnetico del protone presenta una sensibile regolarità su un ampio intervallo delle variabili sperimentali. Il quadrato del fattore di forma magnetico rappresenta il rapporto fra la sezione d'urto di diffusione (dovuta alla magnetizzazione) prodotta da una particella carica di dimensioni finite e quella prodotta sempre da una particella carica ma puntiforme. La dipendenza del fattore di forma dalla quarta potenza della quantità di moto trasferita, cioè da una potenza molto maggiore di quella prevista dai fisici teorici, costituisce il maggior enigma per i fisici che si occupano delle alte energie.

L'ipotesi che la struttura interna del protone e del neutrone, a cui si deve la diffusione anelastica, sia dovuta a entità piú piccole dei nucleoni stessi, sia che questi si trovino nello stato fondamentale sia in uno stato eccitato, è suffragata dai risultati riportati in questo grafico. Le due curve mostrano infatti come la sezione d'urto di diffusione prevista (*curva inferiore*) risulti piú bassa di circa un fattore 40 della sezione d'urto misurata (*curva superiore*). I dati riportati in questo grafico sono rispettivamente relativi a un angolo di diffusione di 6 gradi e a una energia degli elettroni incidenti di 16 GeV.

La diversità delle diffusioni anelastiche prodotte dal protone e dal neutrone costituisce un altro risultato imprevisto negli esperimenti di diffusione. In questo grafico è stato riportato il rapporto fra la sezione d'urto di diffusione anelastica dei due tipi di nucleoni in funzione di una nuova variabile, ω, definita come il rapporto fra i quadrati delle quantità di moto trasferite (*q*) e le differenze fra le energie prima e dopo la diffusione.

singole particelle virtuali, da fotoni in moto nel vuoto; il fotone che li crea può essere sia virtuale che reale. In un certo senso il fotone è, per una breve frazione di tempo, un mesone.

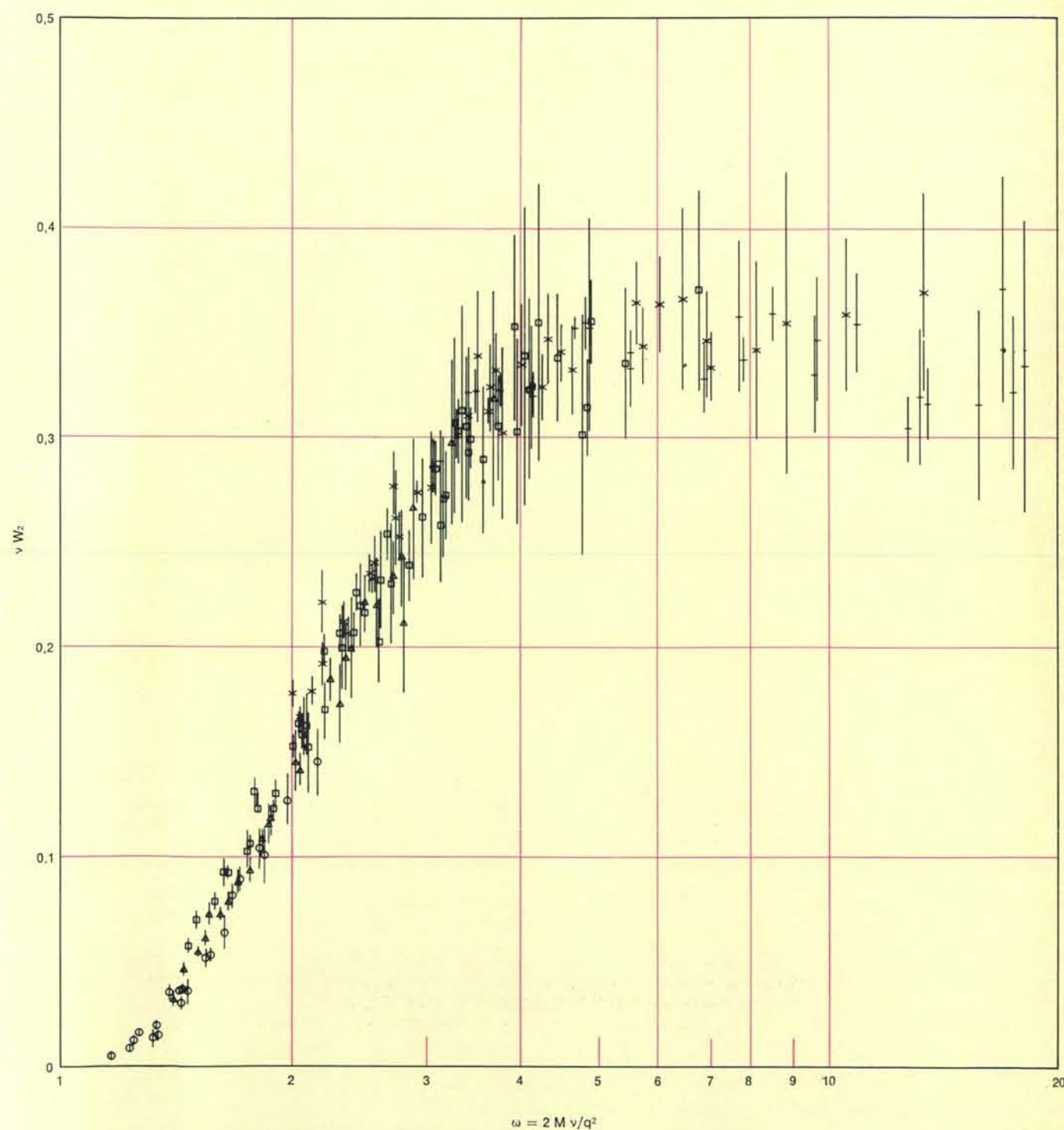
Poiché i mesoni posseggono una massa essi rappresentano un importante fattore nella modificazione dei processi che interessano i fotoni solo negli esperimenti ad altissima energia, quali per esempio gli esperimenti di cui stiamo parlando. Inoltre, in quanto generatori di interazioni forti, i mesoni svolgono un importante ruolo nelle interazioni fra i fotoni reali, di altissima energia e i nucleoni.

Prima che fossero eseguiti questi ultimi esperimenti di diffusione i fisici teorici pensavano di poter prevedere le modalità della partecipazione dei mesoni alla diffusione elastica e anelastica ad alte energie. In particolare essi pensavano che se la diffusione elastica fosse regolata principalmente dai mesoni, la curva del fattore di forma dovrebbe decrescere in funzione dell'inverso del quadrato della quantità di moto trasferita. In realtà la curva decresce in funzione dell'inverso della quarta potenza. È evidente che questo semplice modello non può funzionare.

Correlazioni fra diffusione anelastica ed elastica

In un programma di misure eseguito in collaborazione fra i ricercatori del MIT e dello SLAC sono state determinate per i processi di diffusione anelastica sezioni d'urto notevolmente grandi. Quando si osserva un tipico spettro di diffusione prodotto da elettroni di 10 GeV che collidono con protoni, si nota innanzitutto un netto picco di forma asimmetrica (*si veda la figura a pagina 74*). Il picco rappresenta la diffusione elastica; l'asimmetria dipende dal fatto che gli elettroni possono causare l'emissione di fotoni « molli » (raggi X) di energie diverse, che spostano quindi il picco elastico verso le energie inferiori.

Lo spettro degli elettroni diffusi presenta inoltre due aspetti diversi legati ai processi anelastici. Il primo consiste in una serie di picchi secondari dovuti agli stati eccitati del protone. Essi sono chiamati con il nome di eccitazioni di risonanza, o semplicemente risonanze. I punti di massimo di questi picchi secondari corrispondono agli stati eccitati del protone, identificati in molti esperimenti ad alta energia. Nella diffusione anelastica degli elettroni, sono state identificate quattro risonanze specifiche. La corrispondente ampiezza di questi picchi dipende notevolmente dal valore della quantità di moto trasferita al protone. L'ampiezza dei picchi diminuisce rapidamente all'aumentare della quantità di moto trasferita. La diminuizione è paragonabile alla diminuizione dello stesso picco della diffusione elastica. Da questo concludiamo che le dimensioni radiali degli stati eccitati rappresentati dai picchi secondari sono paragonabili alle dimensioni del protone stesso quando questo si trova nel suo stato fondamentale. Questo implica che sotto molti aspetti nella condizione di stato risonante, cioè nello stato eccitato, il nucleone viene coinvolto con la maggior parte della sua struttura.

Il secondo aspetto dello spettro relativo agli elettroni diffusi anelasticamente è rappresentato da quella parte chiamata « continuo »: cioè da quella distribuzione regolare delle energie degli elettroni diffusi non connessa con i picchi di risonanza. I fisici considerano il continuo come la parte piú interessan-

Questa « curva universale » si ottiene riportando i risultati della diffusione anelastica per un ampio intervallo di angoli di diffusione e di energie iniziali e finali, in funzione della nuova variabile ω introdotta nella figura della pagina precedente. Questa sovrapposizione dei risultati in un'unica curva (una per il protone e una per il neutrone) è consistente con l'ipotesi che la diffusione di elettroni di alta energia sia causata da corpuscoli puntiformi contenuti nei singoli nucleoni. La natura fisica di questi corpuscoli, che sono stati chiamati « partoni », è ancora sconosciuta. Alla sovrapposizione messa in evidenza da questa curva è stato dato il nome di « scaling ». Questo tipo di relazione, che fa intervenire il quadrato della quantità di moto trasferita, è molto comune in tutti i casi in cui viene eseguita la descrizione cinematica della diffusione prodotta da particelle puntiformi.

te ed enigmatica dei risultati ottenuti a Stanford. Se si eseguono le misure ad angoli di diffusione piú grandi o a energie di incidenza maggiori si nota che le risonanze tendono a scomparire mentre il continuo rimane.

Quando è stato promosso il programma di ricerca sulla diffusione anelastica, i teorici credevano che al crescere della quantità di moto trasferita la sezione d'urto relativa al continuo dovesse decrescere piú rapidamente della sezione d'urto relativa alla diffusione elastica. Al contrario i risultati mostrano che per le energie degli elettroni incidenti comprese fra 4,5 e 19 GeV, la sezione di urto di diffusione anelastica è notevolmente simile a quella che si sarebbe avuta per bersagli puntiformi (*si veda la figura qui sotto*). Le migliori previsioni antecedenti gli esperimenti poi eseguiti risultano inferiori di circa un fattore 40 ai risultati degli esperimenti stessi (*si veda la figura in alto a pagina 76*). In altri spettri la differenza risulta ancora maggiore. Una possibile conclusione potrebbe essere che la struttura interna, che dà origine alla diffusione anelastica, sia costituita da entità molto piú piccole del nucleone, sia nello stato fondamentale sia nello stato eccitato.

Una ulteriore conferma che la sezione d'urto di diffusione anelastica sia da attribuirsi a bersagli puntiformi può essere ottenuta dalla osservazione del presente grafico, che fornisce il rapporto fra la diffusione alla Rosenbluth e la diffusione alla Mott, sia nel caso di diffusione elettronica elastica (*curva colorata*), sia per tre diverse regioni dello spettro della diffusione anelastica (*curve nere*). Prima che fossero ottenuti questi risultati, si era ipotizzato che la sezione d'urto relativa al continuo-anelastico presentasse una decrescenza rapida analogamente a quanto avviene per la sezione d'urto elastica.

Il modello a partoni

Richard P. Feynman dell'Istituto di Tecnologia della California (Caltech) sta sviluppando un modello teorico del nucleone che è in grado di spiegare i risultati della diffusione anelastica. Egli, ai costituenti sconosciuti del protone e del neutrone che diffondono anelasticamente gli elettroni di alta energia, ha dato il nome di « partoni ». Feynman suppone che i partoni siano particelle puntiformi. Assieme ad altri ha esaminato la possibilità che i partoni appartengano a una qualche specie delle particelle subnucleari già note. I mesoni che contribuiscono alla « nuvola » di carica nucleonica potrebbero essere i candidati piú probabili. Tuttavia, forti evidenze sperimentali mostrano che i partoni, se in definitiva esistono, non presentano le note proprietà dei mesoni.

Si è anche pensato che i partoni possano essere identificati con quelle ipotetiche entità chiamate quark, cioè con quelle strane particelle proposte indipendentemente nel 1964 da Murray Gell-Mann e da George Zweig del Caltech (si veda l'articolo di Victor F. Weisskopf *Le tre spettroscopie*, in « Le Scienze », n. 2, ottobre 1968). I quark, a differenza di tutte le particelle note, posseggono cariche elettriche non intere, presentano cioè valori $+2/3$ o $-1/3$ ($-2/3$ o $+1/3$ per gli antiquark). Gell-Mann e Zweig suggeriscono che i mesoni potrebbero essere costituiti da un quark e da un antiquark. I nucleoni e le altre particelle che presentano proprietà similari (cioè i barioni) potrebbero invece essere costituite da tre quark. A tutt'oggi non è stata osservata nessuna particella reale con una carica non intera a dispetto delle lunghe e continue ricerche. Ciononostante, è possibile ricostruire matematicamente un dettagliato modello nucleonico, sulla base delle informazioni ottenute dalla diffusione anelastica, assumendo arbitrariamente che questi ipotetici partoni presentino le proprietà formalmente attribuite agli ugualmente ipotetici quark.

I modelli concettuali quali per esempio il modello a partoni, rappresentano lo sforzo fatto dai fisici teorici per descrivere la struttura interna del nucleone in accordo con le piú recenti informazioni fornite dagli esperimenti ad alta energia. I teorici tentano di risolvere i problemi matematici che si presentano quando il modello viene impiegato per « prevedere » le proprietà osservate negli esperimenti che sono già stati eseguiti; inoltre essi suggeriscono ulteriori misure particolarmente adatte a verificare la validità del modello. I modelli vengono scartati o perché le difficoltà matematiche non possono essere superate o perché le previsioni non sono in accordo con i risultati sperimentali. La verifica della bontà di un modello può, come è avvenuto per il modello atomico di Rutherford, estendere enormemente gli interessi e gli scopi della ricerca.

Un ulteriore risultato inaspettato consiste nella differenza sostanziale che si osserva tra la diffusione anelastica prodotta da un protone e quella prodotta da un neutrone (*si veda la figura in basso a pag. 76*). Tuttavia si può notare che si possono semplificare notevolmente i risultati della diffusione di elettroni se si introduce una variabile che rappresenta il rapporto fra il quadrato della quantità di moto trasferita e la differenza fra le energie dell'elettrone prima e dopo il processo di diffusione. Se le diverse osservazioni sono rappresentate graficamente in funzione di questo semplice rapporto, i risultati relativi a un largo intervallo di angoli di diffusione e di energie iniziali e finali si possono riunire in un unico grafico per il neutrone e in un altro per il protone (*si veda la figura a pag. 77*). Questa inaspettata analogia può essere spiegata molto semplicemente se si assume che la diffusione venga prodotta dai singoli partoni, dato che questa relazione che comprende il quadrato della quantità di moto trasferita ha la sua naturale origine nella cinematica della diffusione prodotta da particelle puntiformi. Inoltre la differenza fra la diffusione prodotta dal protone e quella prodotta dal neutrone può essere spiegata qualitativamente sulla base delle diverse configurazioni dei tre quark che costituiscono queste due particelle.

Poiché i partoni, qualsiasi cosa essi siano, sono molto legati fra loro, è estremamente difficile determinare le loro proprietà individuali. Paradossalmente il problema sarebbe piú semplice se si avesse a che fare con una nuvola di partoni in moto, in un sistema di riferimento, con una velocità prossima a quella della luce; in questo caso infatti l'intero nucleone sarebbe relativisticamente contratto e assumerebbe la forma di un disco piatto. In questo caso il fotone virtuale, al quale è associata la forza elettromagnetica esercitata dall'elettrone diffuso, interagirebbe con uno solo dei partoni; il partone (a causa della dilatazione relativistica dei tempi) esisterebbe come particella libera per un tempo sufficientemente lungo e questo potrebbe permettere la determinazione delle sue caratteristiche individuali. Per questo l'analisi teorica di eventi in un sistema di riferimento in rapido movimento può essere eseguita con una maggiore facilità ed essere successivamente riferita a un sistema di riferimento in quiete. In questo modo sarebbe possibile il confronto fra teoria e risultati sperimentali. Per quanto il modello a partoni sia qualitativamente soddisfacente per spiegare i risultati della diffusione, le sue previsioni quantitative non sono del tutto riproducibili. È chiaro che sono necessarie ulteriori informazioni sperimentali e ulteriori studi teorici.

Anche se incompleto, il modello a partoni è stato già impiegato per interpretare i risultati sperimentali di reazioni prodotte da altre particelle, e ha dato lo spunto per altri esperimenti attualmente in fase di progetto. Presso il Centro di ricerche nucleari di Frascati, è stato fatto collidere un intenso fascio di elettroni di alta energia, circolanti in un anello di accumulazione, con un fascio di positoni circolanti in senso opposto. Un certo numero di positoni e di elettroni interagiscono e si annichilano fra loro, dando frequentemente origine a due o piú pioni. La sezione d'urto di annichilazione e di produzione pionica risulta molto maggiore del previsto. La annichilazione elettrone-positone e la « intensa » diffusione anelastica di elettroni osservata a Stanford sono fenomeni direttamente correlati; possono infatti essere considerate come reazioni l'una inversa dell'altra. Per questo la elevata sezione d'urto misurata a Frascati giustifica e conferma gli alti valori di sezione d'urto di diffusione ottenuti a Stanford.

Attualmente sono allo studio molte altre esperienze. In una di queste si vogliono confrontare le diffusioni subite da neutrini e da antineutrini, che dovrebbero presentare sezioni d'urto uguali. Con un'altra serie di misure si vogliono studiare i processi di annichilazione positone-elettrone a valori di energia cosí elevati da permettere la produzione oltre che di pioni anche di protoni e di antiprotoni. (Questa reazione dovrebbe presentare sezioni d'urto analoghe a quelle prodotte da particelle puntiformi).

I risultati relativi alla diffusione elettronica, ottenuti presso l'acceleratore di Stanford, hanno dato un nuovo impulso agli studi teorici e sperimentali. È però ancora presto per dire se il modello a partoni potrà condurre alla comprensione della struttura nucleonica o se invece saranno necessarie altre idee del tutto diverse. In ogni caso, è probabile che una completa spiegazione dei fenomeni di diffusione elettronica servirà a chiarire non solo la natura dei costituenti del nucleone, ma anche la natura delle interazioni forti e le caratteristiche delle particelle soggette a tali interazioni.

La fusione con il laser

Le reazioni di fusione esoenergetiche possono essere iniziate e controllate con una certa sicurezza senza un campo magnetico di confinamento mediante focalizzazione di impulsi laser su una pasticca di combustibile

di Moshe J. Lubin e Arthur P. Fraas

La rapida evoluzione negli ultimi dieci anni dei laser impulsati di alta potenza ha messo a disposizione dei ricercatori di numerosi campi un nuovo strumento la cui potenziale utilità ha da poco cominciato a essere esplorata. Tra le applicazioni finora proposte spicca una varietà di schemi il cui scopo è di evidenziare il fatto che il fascio di luce focalizzato di un laser di tale tipo è in grado di riscaldare una piccola quantità di materia a temperature estremamente alte: in alcuni casi fino a piú di 50 milioni di gradi Kelvin! Per i chimici, per esempio, una tale possibilità equivale a dire che le reazioni chimiche possono essere iniziate in meno di un nanosecondo (un miliardesimo di secondo) e possono essere studiate in condizioni prima considerate irraggiungibili.

Per i fisici del plasma questa possibilità senza precedenti di riscaldare mediante laser a impulsi di alta potenza è particolarmente eccitante. In linea di principio essa apre una nuova promettente strada per affrontare l'annoso problema di come produrre le altissime temperature di plasma necessarie per mantenere in condizioni controllate le reazioni termonucleari o di fusione. Recenti esperimenti hanno mostrato che le reazioni esoenergetiche di fusione, che possono portare come risultato finale alla produzione di energia elettrica utile, possono essere iniziate e controllate con una certa sicurezza con un plasma « confinato inerzialmente » ottenuto focalizzando un potente impulso laser su una densa pasticca congelata fatta di una miscela degli isotopi pesanti dell'idrogeno, deuterio (H^2) e trizio (H^3). In questo articolo passeremo in rassegna lo stato attuale delle ricerche sulle applicazioni dei laser di alta potenza alla produzione di energia di fusione. Il punto sull'energia di fusione e i progressi compiuti negli ultimi 20 anni con tecnologie che non ricorrevano al laser per la produzione di tale energia sono stati trattati nella nostra rivista, in particolare nell'articolo di William C. Gough e Bernard J. Eastlund *Le prospettive dell'energia di fusione* (« Le Scienze », n. 32, maggio 1971). Ci basta ricordare che le reazioni di fusione controllata relative a isotopi leggeri come il deuterio e il trizio rappresentano una fonte potenzialmente inesauribile di energia a buon mercato, di alto rendimento e « pulita », in grado di soddisfare il futuro fabbisogno energetico dell'umanità.

Le possibilità dell'energia di fusione, tuttavia, non sono apparse finora tecnologicamente realizzabili. La difficoltà è duplice. Prima di tutto, il problema di riscaldare il combustibile di elementi leggeri a una temperatura sufficientemente alta con la densità necessaria perché cominci a «bruciare» lentamente si è rivelato ben piú difficile del previsto. La temperatura necessaria per iniziare la combustione varia da 50 a 100 milioni di gradi Kelvin, a seconda del combustibile. In secondo luogo, una volta iniziato il processo di ignizione del combustibile, la calda e densa miscela gassosa di ioni deve essere tenuta insieme per un tempo abbastanza lungo, in modo tale che nella combustione si liberi piú energia di quanta ne è stata necessaria per il processo di ignizione. I tentativi di confinare il plasma mediante intensi campi magnetici di varie geometrie non hanno portato finora ad alcuna soluzione pratica. Però nonostante le difficoltà incontrate nel riscaldare una miscela gassosa a tali temperature e nel confinare il gas ionizzato per grandi frazioni di secondo, le possibilità di recupero di tempo e denaro investiti nelle ricerche su questo problema sono cosí grandi che tutti i maggiori stati del mondo hanno in progetto nutriti programmi di ricerca sulla fusione controllata.

Prima di descrivere come i laser di alta potenza possano avere un ruolo determinante nella soluzione del problema, è istruttivo passare brevemente in rassegna 1) lo stato attuale di sviluppo dei laser a impulsi di alta potenza, 2) il meccanismo di interazione dell'energia di un fascio laser con un mezzo denso e 3) i requisiti necessari perché un laser possa iniziare reazioni di fusione controllata.

Un laser è un dispositivo per la produzione o l'amplificazione di un fascio di luce le cui onde sono sia monocromatiche (tutte con la stessa lunghezza d'onda) sia coerenti (tutte in fase). Il fascio di luce emesso da un laser può essere reso perfettamente parallelo e il suo angolo di divergenza è in teoria limitato solo da effetti di diffrazione. In linea di principio un tale fascio potrebbe essere focalizzato in un punto del diametro di soli duecento milionesimi di centimetro. Il laser produce il suo fascio di luce coerente mediante l'interazione di radiazioni elettromagnetiche col mezzo del laser. Portando piú atomi a un livello energetico superiore di quanti ne esistono a un livello inferiore, l'assorbimento della radiazione eccitante produce nel laser una « inversione » della popolazione atomica. L'energia immagazzinata nell'inversione di popolazione è quindi disponibile per amplificare un'onda luminosa viaggiante a una particolare frequenza. Questa frequenza di emissione è di solito diversa dalla banda di frequenza nella quale il mezzo del laser ha assorbito la radiazione eccitante. Per esempio, un laser a rubino assorbe luce di « pompaggio » nella regione verde dello spettro ottico, ma emette luce coerente nella regione rossa.

Il piccolo impulso di luce da amplificare viene prodotto da un oscillatore a laser. Perché un laser possa funzionare come oscillatore è però necessario fornirgli un meccanismo di retroazione. Questo viene realizzato con l'uso di specchi agli estremi del mezzo laser la cui funzione è quella di far viaggiare avanti e indietro le onde nel laser alla frequenza di emissione. In questo modo si instaurano delle oscillazioni alla frequenza di emissione alimentando l'inversione di popolazione fino al completo consumo dell'energia immagazzinata nel mezzo del laser, dopo di che l'emissione cessa. La prima azione laser su un solido fu osservata nel 1960 in una sbarretta di rubino da T.H. Maiman, che lavorava allora presso la Hughes Aircraft Company. Da allora in poi sono stati scoperti in rapida successione numerosi altri materiali a stato solido per laser. Il piú importante tra questi è il vetro crown al bario drogato con neodimio, la cui azione laser è stata rivelata per la prima volta nel 1961 da Elias Snitzer della American Optical Company. Mentre il rubino è un cristallo la cui crescita fino a dimensioni ragionevoli è costosa, il vetro crown al bario è al confronto di bassissimo costo e può essere costruito in grandi dimensioni e quantità. Le dimensioni diventano un fattore importante nei sistemi laser di altissima potenza.

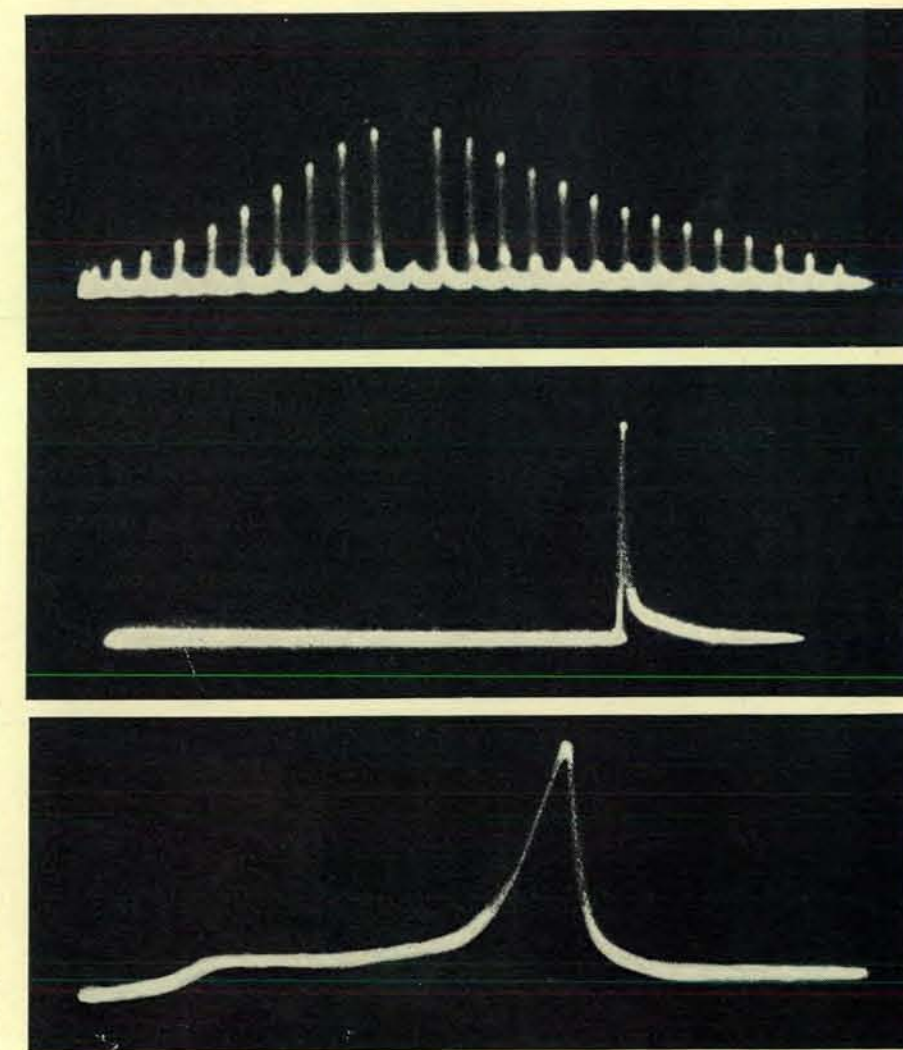
Le superfici estreme della sbarretta laser dalla quale viene emessa la radiazione coerente possono sopportare solo una limitata quantità di energia per centimetro quadrato. Pare accertato che le impurità del materiale laser contribuiscano a un significativo assorbimento della luce laser e ciò porta infine alla rottura del mezzo laser. Per esempio, la tolleranza energetica del vetro al neodimio va da 10 a 100 joule per centimetro quadrato a seconda della durata dell'impulso laser.

Attualmente materiali laser quali il rubino e il vetro al neodimio vengono utilizzati per ottenere i piú alti picchi di potenza in uscita nel funzionamento a impulsi. Questa proprietà è attribuibile principalmente alla maggiore densità degli atomi dei solidi che partecipano all'azione laser. I liquidi che presentano azione laser fanno bene sperare per il futuro nelle applicazioni a impulsi di alta potenza per la stessa ragione; inoltre essi hanno il vantaggio di essere invulnerabili a un danno strutturale permanente. I laser a liquido sotto forma di soluzioni colorate o di vetro liquido sono invece ancora nei primi stadi di sviluppo.

I sistemi laser a stato solido attualmente disponibili sono in grado di produrre piú di 1000 joule di energia in meno di un nanosecondo. Questo flusso di energia coerente appare sotto forma di un fascio altamente collimato con una divergenza totale di meno di un centesimo di grado angolare: un tale fascio di luce che parte con un diametro di quindici centimetri dopo aver percorso un chilometro e mezzo dovrebbe avere un diametro di non piú di 45 centimetri.

Nel caso di un laser a rubino la lunghezza d'onda della luce emessa è di 0,6943 micron. Questa lunghezza d'onda cade nella regione rossa dello spettro ed è perciò visibile. Il vetro al neodimio usato per i laser ad alta potenza emette radiazioni di 1,06 micron, che appartengono alla regione infrarossa dello spettro e non sono perciò direttamente visibili all'occhio umano.

Un sistema laser ad alta potenza (*si veda la figura nelle due pagine seguenti*) è costituito da un oscillatore nel quale si forma un impulso laser a bassa energia, opportunamente formato nello spazio e nel tempo. Questo impulso è un pacchetto contenente un fiotto di radiazioni elettromagnetiche che oscillano alla frequenza del laser. Vi sono molti modi per « modellare » un impulso allo scopo di adattarlo a una particolare applicazione. Impulsi della durata dell'ordine del nanosecondo o poco piú si possono facilmente produrre « tagliando » una parte opportuna di un impulso piú lungo con un veloce otturatore elettroottico. L'otturatore, comandato da un impulso elettrico di alta tensione, è sincronizzato con l'arrivo dell'impulso piú lungo mediante il controllo dell'intensità della radiazione laser in funzione del tempo. Una tale sincronizzazione richiede interruzioni di alte tensioni in tempi dell'ordine del nanosecondo.

Con questa tecnica non si riesce a produrre impulsi di minore durata. Si usa invece comunemente una tecnica conosciuta col nome di *mode-loking* (« modulazione a scatto ») (*si vedano le figure qui sotto*). Quando l'energia viene immagazzinata nella cavità laser per poi essere emessa, dallo specchio semi-

Il taglio di un impulso laser è illustrato in questa sequenza di tracce oscilloscopiche. Un treno di brevissimi impulsi « modulati a scatto » viene dapprima prodotto nella sezione dell'oscillatore del sistema laser (*in alto*). Dal treno viene poi estratto un singolo impulso (mancante nella traccia superiore) che viene poi amplificato (*in mezzo*). La larghezza dell'impulso può variare da un miliardesimo (10^{-9}) a un bilionesimo (10^{-12}) di secondo. A tale impulso viene spesso aggiunto un fronte di salita di bassa potenza per vaporizzare la gocciolina prima di riscaldarla con l'impulso principale.

argentato posto di fronte alla cavità esce una successione di tali impulsi modulati. Anche qui si può usare un otturatore elettroottico, ma solo per « lasciar passare » o selezionare un impulso laser con una durata molto minore del tempo di apertura dell'otturatore. In questo modo si possono lasciar passare uno o piú impulsi. La durata dell'impulso laser è qui determinata unicamente dai componenti della cavità laser. Con questa tecnica si possono ottenere impulsi di durata variabile da un miliardesimo a un bilionesimo di secondo.

Una volta che l'impulso laser opportunamente formato esce dal laser, esso viene diretto su una successione di amplificatori laser ciascuno dei quali aggiunge energia all'impulso originale. Gli amplificatori hanno solitamente aperture sempre piú grandi, in modo che i materiali laser possano resistere senza danni all'aumento di energia. Il processo di amplificazione stesso può essere chiamato in causa per cambiare la forma dell'impulso di entrata. Se l'intensità dell'impulso incidente è sufficientemente elevata, la maggior parte dell'energia dell'amplificatore viene concentrata nel fronte di salita dell'impulso, facendo diminuire in tal modo la larghezza dell'impulso.

Gli atomi della sbarretta di un laser di vetro al neodimio vengono eccitati mediante assorbimento di luce da lampade a flash esterne. Poiché la luce entra nella sbarretta dall'esterno, saranno portati a uno stato energetico eccitato piú atomi dei bordi esterni della sbarretta che dell'interno. Questa situazione sfavorevole diventa intollerabile per sbarrette di piú di cinque centimetri di diametro. Inoltre le sbarrette molto grandi si raffreddano con difficoltà tra i successivi impulsi. Una soluzione al problema di eccitare stadi di amplificazione di grande apertura è l'amplificatore a disco (*si veda l'illustrazione a pagina 84*). Questo tipo di amplificatore è costituito da grandi dischi di vetro al neodimio che vengono orientati rispetto al fascio laser incidente per rendere minime le perdite per riflessione. Nello stesso tempo i dischi hanno una grande sezione trasversale di minimo spessore esposta all'eccitazione delle lampade a flash, il che produce una eccitazione uniforme. Uno di tali amplificatori con un'apertura di quindici centimetri è ora in grado di raggiungere una potenza di uscita di 1000 joule; 100 amplificatori di questo tipo potrebbero essere disposti in parallelo per ottenere uscite di 100 000 joule.

Sono disponibili numerose sorgenti per produrre tali fiotti di energia in breve tempo. Nessuna uguaglia tuttavia il laser per quanto riguarda la facilità con cui l'uscita può essere concentrata in un piccolo volume. È proprio questa capacità di focalizzare radiazioni di alta potenza, mediante l'uso di elementi ottici quali lenti e specchi, che rende il laser particolarmente adatto allo scopo di iniziare la fusione. La potenza di picco in un tal fascio prima della focalizzazione è superiore a 10^{13} watt, mentre, quando il fascio viene focalizzato, è possibile ottenere densità di potenza superiori a 10^{17} watt al centimetro quadrato. La focalizzazione di energia ha come risultato un campo elettrico locale nelle immediate vicinanze del fuoco dell'ordine di 10^{10} volt al centimetro! È difficile rendersi conto della grandezza di tali enormi intensità di campi elettrici; un'idea di tale intensità la possiamo avere pensando che un elettrone libero accelerato da un campo di 10^{10} volt al centimetro raggiungerebbe una velocità confrontabile con quella della luce in una distanza pari a una frazione di una lunghezza d'onda ottica.

Il tipico sistema laser ad alta potenza illustrato in queste due pagine fu progettato per un esperimento di riscaldamento del plasma. Il sistema è costituito da una serie di componenti di vetro crown al bario drogato con neodimio e comincia con un oscillatore laser (*estrema sinistra*) nel quale viene formato un impulso laser di bassa energia. L'impulso viene poi tagliato, cioè formato sia nel tempo sia nello spazio per adattarlo a una particolare applicazione, prima di essere amplificato da sbarrette laser addizionali e alla fine da un grande amplificatore a disco. Un'opportuna combinazione di prismi manda l'impulso tre volte nell'amplificatore a disco. Il fascio risultante, di circa 15 centimetri di diametro, viene infine focalizzato su una gocciolina di deuterio in una camera a vuoto (*estrema destra*) per produrre un plasma caldo, o gas ionizzato, nel quale hanno luogo le reazioni di fusione. Le reazioni di fusione vengono studiate mediante i neutroni energetici da esse emessi. La potenza di picco comunemente ottenibile con un tale sistema prima della focalizzazione supera i 10^{13} watt. Quando viene focalizzato è possibile raggiungere densità di potenza di 10^{17} watt al centimetro quadrato.

È ben noto che i combustibili per la fusione nucleare, quali il deuterio e il trizio, liberano piú energia per nucleone (neutrone o protone) dei combustibili della fissione nucleare. Sfortunatamente questi isotopi dell'idrogeno bruciano con buon rendimento solo ad alta temperatura. Al di sotto dei 10 milioni di gradi Kelvin la combustione è troppo lenta per poter essere utile in un processo di produzione di energia qui sulla Terra. Il Sole può bruciare idrogeno in modo efficiente a temperature inferiori a causa del suo grande volume. Nelle dimensioni limitate di un laboratorio, per il plasma si devono prendere in considerazione isotopi che brucino a temperature superiori a 50 milioni di gradi Kelvin.

La tradizionale tecnica di avvicinamento alla fusione controllata ha dovuto iniziare dalla produzione di un plasma diluito e piuttosto freddo. Il plasma viene successivamente intrappolato in un campo magnetico di confinamento. Si passa poi al riscaldamento del plasma a temperatura superiore tenendolo fermo con un campo magnetico che viene reso proporzionalmente piú intenso. Le nostre attuali conoscenze ci permettono solo di riscaldare il plasma con cautela in modo tale da non sconvolgere il delicato equilibrio tra il gas caldo ionizzato e il campo magnetico. Sebbene le nostre conoscenze sulle forze e sulle condizioni per il controllo di questo equilibrio siano notevolmente aumentate negli ultimi dieci anni non siamo ancora sicuri di quale sia la piú adatta configurazione di campo magnetico necessaria per contenere il plasma caldo abbastanza a lungo perché possa avvenire una apprezzabile combustione.

Quando cominciarono a funzionare sistemi laser a impulsi di alta potenza, una delle prime loro applicazioni nel programma sulla fusione controllata fu la produzione di plasmi estremamente puliti e di buona qualità per riempire i dispositivi esistenti per il confinamento magnetico (*si veda lo schema in alto a pag. 85*). La domanda che venne posta in quel momento fu: « Un plasma prodotto riscaldando col laser una gocciolina di combustibile termonucleare in un campo magnetico può svolgere le funzioni del plasma necessario per un processo continuo quasi stazionario di reazioni di fusione »?

Piú recentemente, con l'aumento della capacità di produrre energia dei sistemi laser a impulsi di alta potenza, è sorta una seconda domanda: « È possibile utilizzare l'intenso fiotto di radiazione coerente focalizzata dal laser per aumentare la temperatura di un opportuno bersaglio deuterio-trizio a un valore tale da produrre una apprezzabile combustione *senza* che sia necessario un campo magnetico di confinamento»?

Per rispondere a queste domande si deve dapprima esaminare in quale modo la radiazione elettromagnetica di alta frequenza interagisce con un plasma denso di elettroni e di ioni positivi. Il campo elettrico oscillante di un fascio di luce può essere assorbito, riflesso, diffuso e/o rifratto da una superficie materiale. Il nostro interesse è principalmente rivolto all'assorbimento di questa radiazione e alla conversione dell'energia elettromagnetica del fascio laser in energia termica del combustibile di fusione. Una volta che sulla superficie dell'isotopo solido dell'idrogeno siano stati prodotti alcuni elettroni non legati, questi elettroni liberi raccolgono rapidamente energia dal campo elettrico oscillante incidente. La loro energia viene poi trasferita agli ioni del mezzo del plasma mediante interazioni attrattive a lunga distanza. Cosí anche gli ioni piú pesanti « sentono » l'effetto del fascio laser attraverso la mediazione degli elettroni oscillanti. Il trasporto di questa energia dagli elettroni agli ioni può essere molto piú lento del riscaldamento degli elettroni da parte del fascio laser. In effetti si dubita che in un plasma libero di deuterio-trizio i trasferimenti efficaci di energia possano avvenire in tempi di ordine minore di 10^{-11} secondi. Questo processo di assorbimento è chiamato assorbimento per urto.

Se la radiazione incidente deve essere assorbita con efficacia, il plasma dovrebbe essere opaco per essa. D'altra parte è desiderabile che la radiazione laser non venga dispersa per riflessione. In termini piú quantitativi, è auspicabile che l'onda luminosa incidente ceda tutta la sua energia agli elettroni nelle prime poche collisioni con gli ioni vicini. Questo modello ci permette di valutare quale debba essere lo spessore di un plasma denso per assorbire tutta la luce laser incidente. Quando un pla-

L'amplificatore a disco è in grado di produrre un impulso di energia in uscita superiore a 1000 joule da un'apertura di 15 centimetri di diametro. I dischi laser di vetro al neodimio sono montati inclinati rispetto al fascio laser incidente in modo da rendere minima la riflessione nell'istante in cui essi vengono eccitati intensamente sulle facce da lampade al flash disposte nel riflettore da entrambi i lati dei dischi. Un centinaio di amplificatori di questo tipo disposti in parallelo potrebbe in linea di principio in base ai calcoli teorici, far raggiungere all'uscita un'energia addirittura superiore a 100 000 joule.

sma è circondato dal vuoto, la densità di elettroni e di ioni cresce rapidamente da zero sulla superficie esterna fino a un massimo nel centro del plasma. La maggior parte della radiazione laser viene assorbita in questo strato di densità crescente (*si veda la figura in basso nella pagina a fronte*). Per un completo assorbimento in un plasma denso alle temperature termonucleari questo strato richiede uno spessore di soli 0,012 centimetri.

La risposta alla prima domanda è stata data da lavori teorici e sperimentali eseguiti in varie parti del mondo. Negli Stati Uniti un gruppo guidato da Alan F. Haught degli United Aircraft Research Laboratories è sul punto di dimostrare che i plasmi prodotti mediante il riscaldamento di piccole pasticche con i laser possono essere confinati nelle esistenti geometrie di campi magnetici in un ampio intervallo di temperature. Il plasma risultante, quasi stazionario, sebbene non sufficientemente caldo per significative reazioni di fusione controllata, può poi essere arricchito e riscaldato a temperature superiori con altri metodi. In tal modo plasmi puliti prodotti con i laser possono essere il primo passo verso il raggiungimento delle condizioni di fusione controllata. Poiché il plasma del laser è solo il catalizzatore per il raggiungimento di un processo di fusione in stato stazionario, il basso rendimento del sistema laser non ha alcuna importanza particolare. La seconda domanda – relativa alla realizzazione del riscaldamento fino alle condizioni termonucleari di un piccolo plasma denso mediante il laser senza che sia necessario un campo magnetico di confinamento – accentra un interesse sempre crescente. Le condizioni necessarie per ottenere questo risultato possono essere meglio comprese analizzando le varie fasi del processo di riscaldamento di un plasma prodotto con il laser. Nella disposizione sperimentale tipica la radiazione di un laser a impulsi viene focalizzata su una piccola pasticca di combustibile solido deuterio-trizio. La pasticca di combustibile solido viene trasformata in un plasma denso dal fronte di salita dell'impulso laser incidente, che è di durata sufficiente per vaporizzare (ma non riscaldare apprezzabilmente) la gocciolina. Una volta che la gocciolina è stata ridotta in forma gassosa viene applicato l'impulso principale per riscaldarla.

Il plasma non rimane stazionario mentre viene riscaldato. Una dettagliata analisi del processo di assorbimento mostra che la dinamica del processo di espansione del plasma riscaldato dalla radiazione incidente è di vitale importanza per la dipendenza dal tempo del processo di assorbimento. Una tipica rappresentazione grafica del processo di assorbimento in funzione del tempo per una gocciolina inizialmente solida (*si vedano i diagrammi a pagina 86*) mostra l'esistenza di un istante al di là del quale il plasma si è espanso fino a diventare trasparente alla luce laser incidente. La durata dell'energia laser utilizzabile per riscaldare piccoli plasmi densi in espansione libera è limitata da due tempi caratteristici. Il primo è il tempo necessario perché il plasma si espanda fino al punto in cui è troppo rarefatto per poter continuare ad assorbire una apprezzabile frazione dell'energia laser incidente (alcuni nanosecondi). Il secondo è il tempo associato al trasferimento di energia elettroni-ioni. Come abbiamo già detto, questo tempo è dell'ordine di grandezza di 10^{-11} secondi. Quindi un impulso laser della durata di 10^{-10} secondi, compresa tra questi due tempi caratteristici, è ben adatto a questo scopo. L'energia laser immagazzinata con alto rendimento nel plasma appare inizialmente sotto forma di energia termica e questa energia, associata con l'energia di fusione, viene alla fine suddivisa tra i prodotti della reazione di fusione, perdite per radiazione e energia per l'espansione.

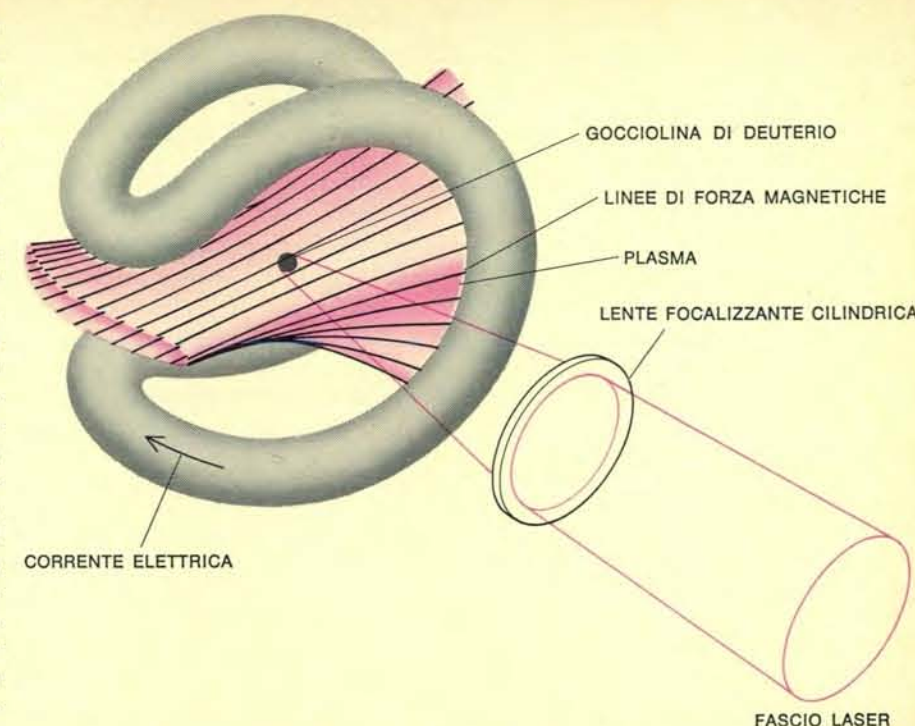
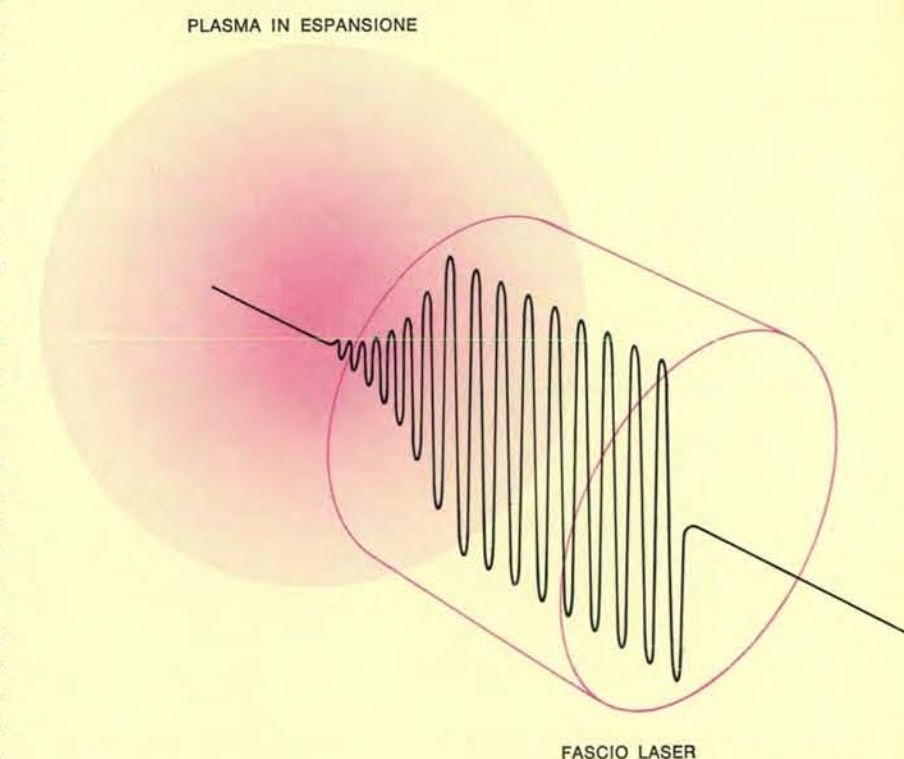
In linea di principio si potrebbe diminuire la quantità di energia utilizzata per l'espansione mediante un intenso campo magnetico di confinamento. Ciò richiederebbe però campi magnetici in grado di esercitare pressioni di milioni di atmosfere. Campi di tale intensità non sono attualmente realizzabili.

Una volta note in funzione del tempo la densità e la temperatura degli ioni, si determina facilmente la resa termonucleare. Calcoli rigorosi su tali tipiche configurazioni indicano che è possibile ottenere un rapporto tra energia liberata ed energia all'ingresso maggiore di uno per energie di ingresso tra 10^5 e 10^6 joule. Le ipotesi sulle quali sono basati questi calcoli sono attualmente in fase di studio sperimentale in un gran numero di laboratori.

I neutroni prodotti nelle reazioni di fusione sono il « termometro » ideale per una valutazione del successo di un particolare esperimento di riscaldamento ad alta potenza. Per esempio, in ogni reazione deuterio-trizio viene prodotto un neutrone con un'energia di 14,1 MeV. Il numero di tali reazioni dipende fortemente dalla temperatura. Dei neutroni prodotti nell'interazione tra una luce laser e un plasma ha per la prima volta dato un resoconto N.G. Basov dell'Istituto Lebedev di Mosca nel 1963. Il lavoro dei russi è stato seguito da esperimenti con neutroni ottenuti con laser ad alta potenza e plasma al Laboratorio Limeil in Francia e in diversi luoghi degli USA, e precisamente all'Università di California a Livermore, al Laboratorio nazionale di Los Alamos, alla Sandia Corporation e all'Università di Rochester. Altri laboratori dei vari paesi del mondo si sono serviti di quei risultati iniziali per migliorare le loro attrezzature. Sono state misurate, usando bersagli di deuterio bombardati con energie del fascio laser tra i 50 e i 250 joule, rese totali di neutroni superiori a 10^5 ottenuti dalla reazione deuterio-deuterio. L'interpretazione di questi risultati indica un assorbimento lievemente migliore di quello che prevede la semplice teoria delle collisioni.

È di particolare interesse lo studio dettagliato dell'assorbimento della radiazione incidente. Vi sono molti importanti contributi all'assorbimento che possono essere piú significativi del semplice processo di collisione descritto prima e che possono quindi influenzare la liberazione di energia. La natura sembra tuttavia favorire un plasma denso in espansione libera, poiché la maggior parte di questi particolari processi di assorbimento fa aumentare l'efficacia del fascio laser incidente.

Attualmente laser con impulsi opportunamente formati possono liberare 10^3 joule, mentre ne sono necessari almeno 10^5 per realizzare una quantità apprezzabile di fusione. Questa energia laser potrebbe essere ottenuta anche usando l'attuale tecnologia, ma non esiste ancora nessun impianto di questo tipo in funzione; sebbene i risultati sperimentali siano pochi, essi appaiono giustificare un certo ottimismo.

L'energia liberata da una pasticca di combustibile denso delle dimensioni di un centimetro mediante la focalizzazione di radiazione laser può essere convenientemente utilizzata per un gran numero di differenti applicazioni di conversione dell'energia. Per esempio, una centrale che produca energia elettrica da reazioni termonucleari controllate mediante la vaporizzazione periodica di

Il primo uso dei laser ad alta potenza nel programma sulla fusione controllata è stato per la produzione di un plasma che riempisse i dispositivi esistenti per il confinamento magnetico come quello indicato in questo disegno schematico. Una corrente elettrica che attraversa un avvolgimento a « cucitura da palla da baseball » induce un campo magnetico la cui densità di energia è minima al centro e cresce verso l'esterno in tutte le direzioni, creando in tal modo un « pozzo » magnetico per il confinamento del plasma generato riscaldando col laser una gocciolina di combustibile posta nel centro. In un esperimento eseguito nel 1968 all'Università di Rochester un dispositivo di questo tipo venne usato per la produzione di un plasma confinato a una temperatura di 10 milioni di gradi Kelvin con una densità di 10^{13} ioni al centimetro cubo.

L'impulso laser viene assorbito con grande efficacia da una sfera di plasma denso in espansione; la maggior parte dell'assorbimento ha luogo in un sottile strato superficiale di plasma dello spessore di soli due centesimi di centimetro circa. Mentre l'energia del fascio laser viene assorbita dal plasma, l'ampiezza del campo elettrico oscillante (*curva nera*) diminuisce rapidamente all'interno di questo sottilissimo strato superficiale.

pasticche di deuterio-trizio presenterebbe un aspetto ben diverso da una configurazione progettata per essere usata, per esempio, nella propulsione spaziale. Per chiarire le idee consideriamo ora le possibilità di realizzazione di questo schema generale per un impianto di energia termonucleare controllata.

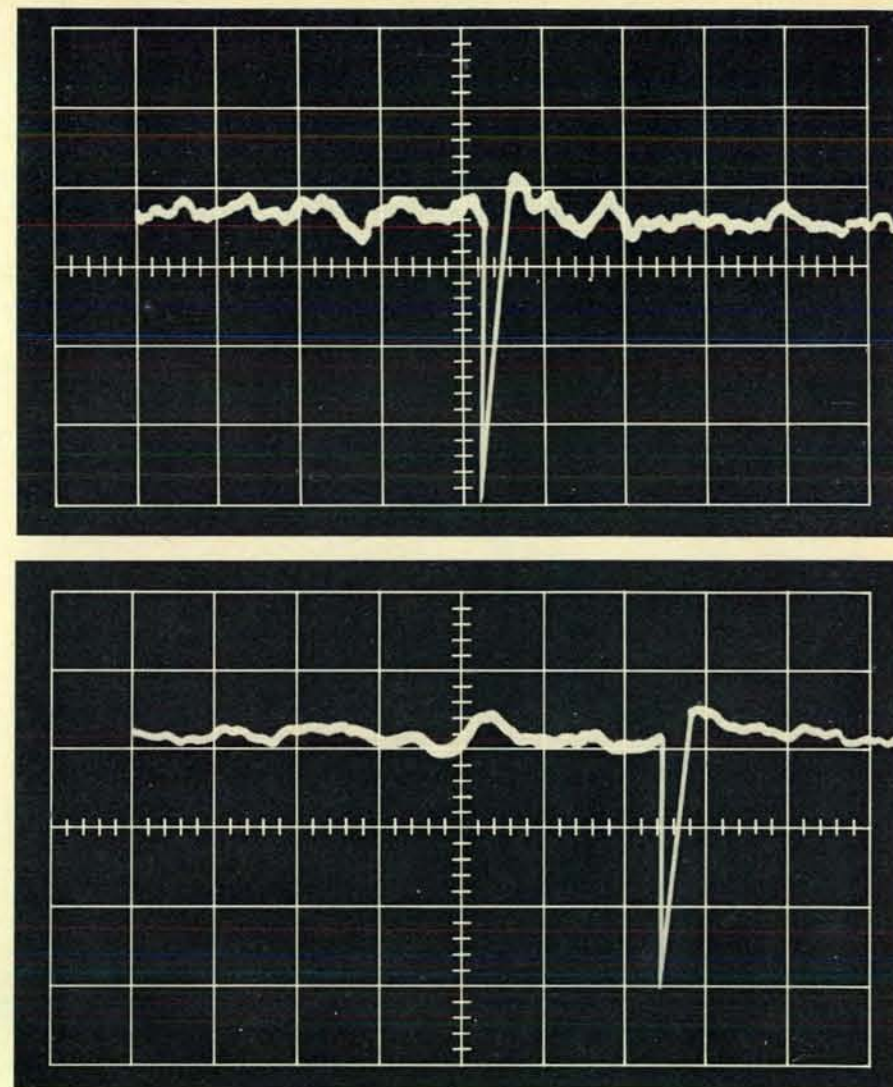
Un metodo per la conversione della energia di fusione di una pasticca di deuterio-trizio innescata col laser in energia elettrica è stato sviluppato fin dal 1969 presso il Laboratorio nazionale di Oak Ridge insieme agli studi sulla realizzabilità dell'energia di fusione che erano già stati iniziati fin dal 1967. Il metodo di Oak Ridge per lo sfruttamento dell'energia di fusione consiste nell'assorbire l'energia di fusione in una vasca di litio, che fornisce a sua volta energia sotto forma di calore a un ciclo termodinamico.

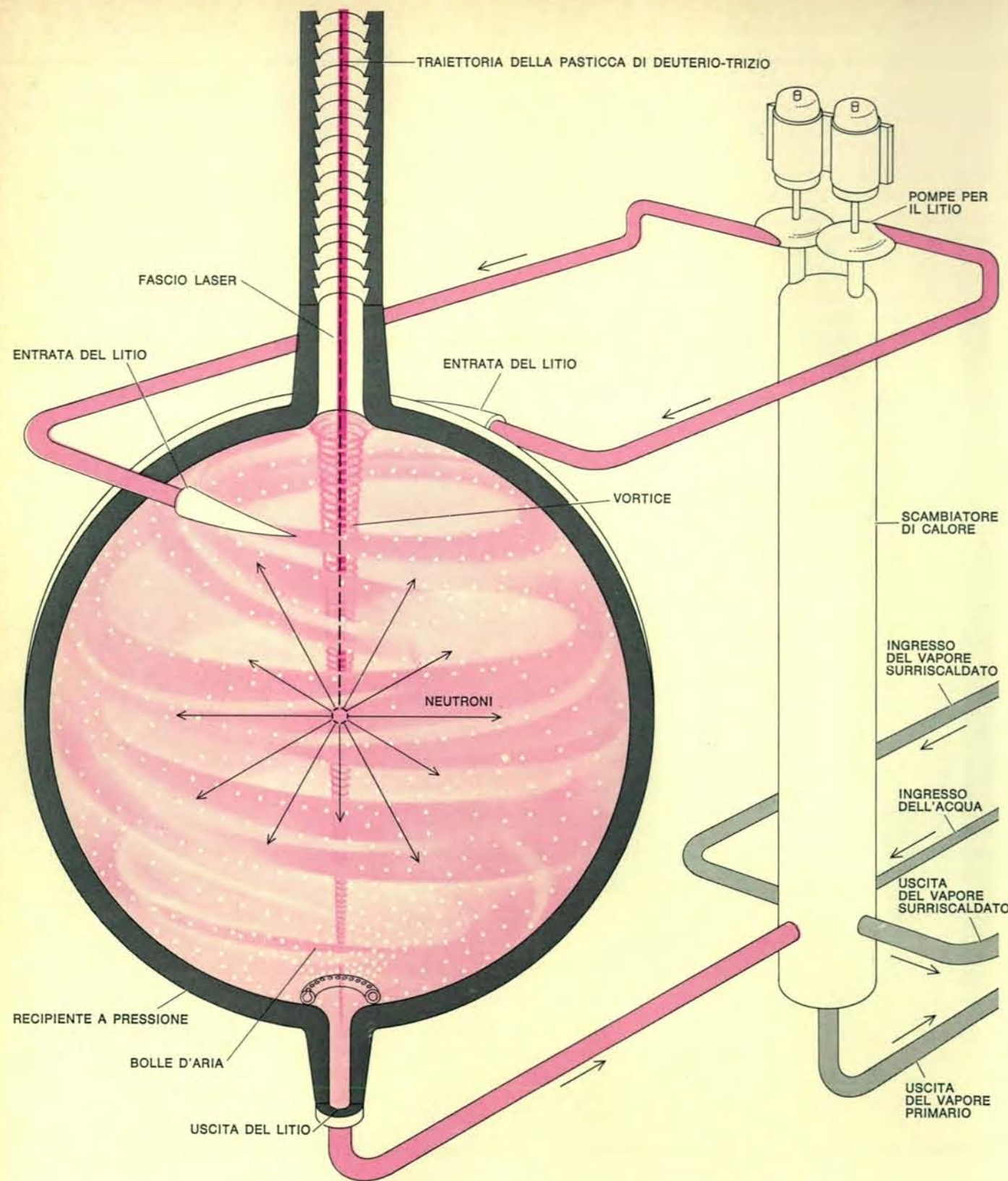
La vasca di litio dovrebbe essere contenuta in un recipiente a pressione cilindrico o sferico di diametro da 3 a 4,5 metri e dovrebbe essere fatta ruotare a una velocità sufficientemente alta da formare un vortice libero attorno al suo asse verticale (*si veda lo schema a pagina 88*). Regolando opportunamente la distribuzione di velocità della rotazione dovrebbe essere possibile ottenere una cavità centrale con un diametro pressocché uniforme di circa cinque centimetri attraverso la regione che si estende dalla cima del recipiente fino a ben al di sotto del piano mediano. Una pasticca ghiacciata di deuterio-trizio dovrebbe essere iniettata nel centro della cavità a vortice e dovrebbe essere innescata con un impulso laser appena raggiunto il piano mediano. L'energia immagazzinata nel litio sotto forma di calore dovrebbe essere rimossa estraendo il litio dal fondo del recipiente a pressione, facendolo passare in scambiatori di calore e quindi reimmettendolo nel recipiente a pressione. Il processo dovrebbe ripetersi circa una volta ogni 10 secondi. La grande inerzia termica del circuito del litio dovrebbe riuscire a mantenere un flusso di calore praticamente costante nel ciclo termodinamico.

Una tipica rappresentazione grafica del processo di assorbimento in funzione del tempo per una pasticca di combustibile di deuterio riscaldata col laser è illustrata in questi quattro grafici. La pasticca, che in origine aveva un diametro di un decimo di millimetro, veniva riscaldata da un impulso laser tagliato della durata di 10^{-10} secondi e con una potenza di picco di 10^{13} watt. La curva più marcata nel primo grafico rappresenta la temperatura degli elettroni, quella più leggera la temperatura degli ioni. L'impulso incidente veniva assorbito unicamente con processi non relativistici e quindi la produzione risultante di neutroni, che rappresenta una importante misura del successo della combustione, costituisce una valida stima.

L'estrazione del litio dal fondo del recipiente a pressione dovrebbe favorire la stabilizzazione del vortice. Dopo essere passato attraverso gli scambiatori di calore il litio dovrebbe ritornare mediante pompe a ugelli tangenziali disposti sul perimetro del recipiente a pressione, mantenendo perciò il vortice desiderato.

Una obiezione evidente a questo progetto è che in tale modo non si riuscirebbero a produrre quantità apprezzabili di energia, poiché la quantità di energia che si potrebbe liberare in ogni esplosione dovrebbe essere mantenuta bassa per evitare che scoppi il recipiente a pressione. Un'analisi relativamente semplice dimostra che le cose stanno realmente cosí a meno che non si prendano provvedimenti per attenuare l'onda d'urto prima che colpisca le pareti del recipiente a pressione. In effetti, si possono prendere diverse precauzioni per far degradare piú rapidamente l'energia nell'onda d'urto. Una dovrebbe essere quella di introdurre un gran numero di bolle di gas per ridurre la densità media del liquido di circa il 5 %. Quando l'onda d'urto attraversa una bolla di gas una parte della sua energia viene usata per espellere il liquido dalla superficie interna della bolla e per proiettarlo nel vuoto fino a colpire le pareti della bolla dall'altra parte, dove la maggior parte dell'energia cinetica del materiale proiettato dovrebbe essere trasformata in energia termica. È difficile valutare fino a che punto dovrebbe attenuarsi l'onda d'urto principale in questo modo, ma semplici prove indicano che con una densità di bolle sufficientemente alta si dovrebbe poter aumentare l'energia liberata in ogni esplosione di un fattore 10 rispetto al valore corrispondente per una vasca di litio senza bolle.

Le operazioni ora descritte sono progettate allo scopo di ridurre l'intensità dell'onda d'urto prima che colpisca le pareti del recipiente a pressione, ma non dovrebbero avere alcun effetto sull'onda d'urto che procede verso l'alto nella cavità del litio liquido verso il sistema di iniezione del laser e della pasticca. Fortunatamente la bassa densità del vapore di litio fa diminuire l'intensità dell'onda d'urto. Inoltre i tubi per l'iniezione della pasticca e per il fascio laser si possono fare molto lunghi, con profilo delle pareti tale da rompere l'onda d'urto normale in un gran numero di onde d'urto oblique, dissipando in tal modo la maggior parte della sua energia.

I fiotti di neutroni che arrivano successivamente su due rivelatori posti a 225 centimetri di distanza vengono segnalati dalle nette deflessioni di queste due tracce oscilloscopiche. La differenza tra gli istanti di arrivo è una misura dell'energia dei neutroni e indica che questi particolari neutroni sono stati prodotti da reazioni deuterio-deuterio nel plasma sperimentale. Un esperimento analogo eseguito all'interno di un plasma di idrogeno puro non ha portato invece ad alcuna produzione rilevabile di neutroni.

Cerchiamo ora di analizzare in dettaglio la successione degli eventi nel corso di un ciclo di espansione, cominciando dall'iniezione nel vortice della pasticca di deuterio-trizio. Mentre la pasticca percorre la sua traiettoria dovrebbe venirne vaporizzato circa il 10 % in conseguenza dell'assorbimento di calore irraggiato verso di essa dal litio caldo. La restante parte di pasticca dovrebbe essere vaporizzata e ionizzata quando essa viene colpita dal fascio laser. Alcuni ioni di deuterio e di trizio dovrebbero subire la fusione e l'energia di fusione liberata dovrebbe far aumentare la temperatura della restante parte della pasticca in modo che circa il 5 % degli ioni dovrebbero subire la fusione prima che l'espansione del plasma faccia finire la reazione. Circa il 75 % dell'energia della reazione di fusione dovrebbe apparire sotto forma di neutroni energetici e dovrebbe poi essere assorbita quando quei neutroni vengono rallentati nella regione piú interna del litio. L'energia assorbita dovrebbe far aumentare la temperatura della regione centrale del litio di qualche centinaio di gradi in circa un milionesimo di secondo e dovrebbe provocare l'espansione del litio; tale espansione, però, dovrebbe essere largamente compensata dalla contrazione delle bolle in modo che non dovrebbe aversi alcuna

Questo reattore a fusione è stato progettato dai ricercatori del Laboratorio nazionale di Oak Ridge come strumento per la conversione dell'energia di fusione ottenuta da pasticche di deuterio-trizio portate all'ignizione col laser in energia elettrica utile. L'energia di fusione dovrebbe essere assorbita in una vasca di litio contenuta in un recipiente a pressione sferico o cilindrico da 3 a 4,5 metri di diametro. Un vortice libero dovrebbe essere mantenuto stazionario attorno a un asse verticale mettendo in rotazione il litio a velocità sufficientemente alta. Una pasticca congelata di deuterio-trizio dovrebbe essere iniettata nel centro della cavità del vortice e in seguito venir portata all'ignizione con un impulso laser una volta raggiunto il piano mediano. L'energia immagazzinata nel litio come calore sotto forma di neutroni energetici (*frecce nere*) dovrebbe essere asportata estraendo il litio dal fondo del recipiente, facendolo passare attraverso scambiatori di calore e facendolo poi ritornare attraverso pompe centrifughe a fori tangenziali praticati lungo il perimetro del recipiente. In questo particolare progetto, l'onda d'urto creata dall'esplosione della pasticca dovrebbe venire attenuata per proteggere il recipiente a pressione introducendo nella vasca di litio, attraverso un anello forato posto sul fondo del recipiente, delle bolle di gas che assorbano l'urto. I sistemi di iniezione della pasticca e del laser dovrebbero essere protetti allungando i tubi per l'iniezione della pasticca e dotandone le pareti di un profilo che spezzi l'onda d'urto normale in tante onde d'urto oblique, attenuando così l'onda primaria.

apprezzabile onda d'urto. Circa il 25% dell'energia liberata dalla reazione di fusione dovrebbe apparire sotto forma di particelle alfa, le quali dovrebbero, a loro volta, causare il riscaldamento degli ioni e, di conseguenza, l'emissione di raggi X. L'energia dei raggi X dovrebbe venire assorbita nei primi millimetri dello spessore di litio, provocando la vaporizzazione di questo sottile strato. La vaporizzazione del litio insieme all'espansione del plasma termonucleare dovrebbe provocare una debole onda d'urto nel litio.

La velocità del suono nel litio caldo è di circa 4500 metri al secondo; il liquido dovrebbe muoversi dal centro dell'esplosione verso l'esterno a una minor velocità in modo da formare una cavità sferica. Il liquido spostato dovrebbe muoversi verso la regione superiore della cavità del vortice e questo fatto dovrebbe permettere un buon mescolamento, in particolare nella regione dei tubi per l'iniezione della pasticca di combustibile e del fascio laser. La geometria dei tubi di iniezione, progettata in modo tale da impedire all'onda d'urto di raggiungere le entrate, dovrebbe tuttavia essere in grado di impedire anche al liquido di affluire fino alle entrate.

Il momento angolare del litio nel recipiente a pressione dovrebbe favorire la riformazione del vortice. Inoltre il flusso di litio dovrebbe essere talmente elevato che in ogni ciclo dovrebbe venire sostituito circa il 50 % del litio nel recipiente.

I problemi degli sforzi ai quali dovrebbe essere sottoposto il recipiente a pressione sono stati esaminati da C.V. Chester e da Lawrence Dresner a Oak Ridge. Essi hanno potuto rapidamente giungere alla conclusione che il sistema proposto presenta tante complessità analitiche, specialmente se si tiene conto degli effetti delle bolle di gas nel liquido, da far dubitare che sia possibile sviluppare un modello analitico su basi puramente teoriche. Ciò nondimeno, pare che a tutti i fenomeni coinvolti si possa applicare per lo studio degli effetti dell'urto la legge dei modelli in scala ridotta comunemente usata. Di conseguenza le prove sperimentali sono continuate su modelli piccoli.

Chester e Dresner hanno progettato una serie di prove su modelli usando piccoli recipienti di acciaio e comuni cariche di esplosivo. Le prime prove sono state eseguite col recipiente a pressione in scala riempito d'acqua senza bolle. La quantità di esplosivo è stata gradualmente portata da 1,5 grammi a 10 grammi. La dilatazione del recipiente all'altezza del piano mediano è stata misurata dopo ogni prova con un micrometro e sono stati rappresentati graficamente i risultati ottenuti in funzione del peso di esplosivo. Per ricavare il massimo valore della carica che si poteva far esplodere senza deformare il recipiente al di là del limite di elasticità è stata tracciata una retta che attraversasse la banda dei punti sperimentali e passasse per il punto a dilatazione nulla. Le prove eseguite hanno indicato che il valore trovato era in accordo con quello ricavato analiticamente e che corrispondeva a circa due grammi di esplosivo. In tal modo il modello analitico semplificato fu collaudato e preso in considerazione per recipienti di dimensioni reali.

Una seconda serie di esperimenti fu eseguita con un sottile strato di gomma spugnosa disposto sulla parete interna del recipiente a pressione. Questo smorzò sufficientemente l'onda d'urto da poter consentire di aumentare fino a piú di quattro grammi la quantità di esplosivo da usare senza deformare il recipiente al di là del limite di elasticità.

Una terza serie di esperimenti fu eseguita con un anello forato posto sul fondo del modello del recipiente a pressione. Venne immessa aria nell'anello per formare una cortina anulare di bolle nella regione vicina alla parete del recipiente a pressione. Un misuratore di sforzi montato esternamente al modello nello stesso piano orizzontale dell'esplosivo permetteva di misurare chiaramente lo sforzo del recipiente sia all'inizio, senza bolle d'aria nel recipiente, sia alla fine, con il 5 % di bolle d'aria nel recipiente.

Nell'estrapolare questi risultati a un sistema di formato naturale – un recipiente a pressione 25 volte piú grande, con un diametro di circa 4 metri e uno spessore di pareti di 25 centimetri – l'energia liberata nell'esplosione dovrebbe aumentare di un fattore 25^3, (15 625). Fissando un fattore di sicurezza due, si potrebbero impiegare 15 chilogrammi di esplosivo in un sistema in formato naturale usando acqua senza bolle a temperatura ambiente. Ciò corrisponde alla liberazione di energia di 15×10^5 calorie per esplosione. Se l'esplosione venisse ripetuta con una frequenza di una ogni 10 secondi, la potenza all'uscita dovrebbe essere di 6,3 megawatt.

Tale valore di 6,3 megawatt corrisponde alla potenza dell'onda d'urto in un sistema nel quale si faccia esplodere del TNT (trinitrotoluene o tritolo) in acqua senza vortici o bolle per attenuare l'esplosione. Con il TNT tutta l'energia verrebbe concentrata nell'onda d'urto, mentre in un reattore a fusione circa l'80 % di questa energia dovrebbe apparire sotto forma di calore nel liquido in conseguenza del rallentamento e dell'assorbimento di neutroni veloci. La maggior parte di questa energia andrebbe semplicemente a riscaldare il liquido e a comprimere le bolle anziché concentrarsi nell'onda d'urto; perciò la produzione di potenza del reattore potrebbe essere all'incirca cinque volte superiore a quella che produce l'onda d'urto, e cioè di circa 30 megawatt. Le prove eseguite sul modello indicano che l'effetto di attenuazione delle bolle dovrebbe ridurre l'intensità dell'onda d'urto di almeno un altro fattore 2,5 e che l'uso di un recipiente sferico anziché di uno cilindrico dovrebbe far guadagnare un altro fattore due, consentendo in tal modo di aumentare la potenza in uscita fino a circa 150 megawatt. Ulteriori aumenti di potenza potrebbero essere possibili in conseguenza della differenza di comprimibilità tra l'acqua e il litio ad alta temperatura, ma non si dispone di dati per una valutazione dell'entità di questo effetto.

Il progetto di un impianto di potenza in formato naturale è stato ricavato dalle idee qui esposte con le limitazioni imposte da considerazioni fondamentali sul progetto. Per esempio, pur essendo il deuterio di basso costo e facilmente reperibile, il trizio non lo è. Di conseguenza, perché un reattore a fusione a deuterio-trizio sia economicamente conveniente, esso deve produrre trizio. Ciò si può ottenere poiché i neutroni energetici hanno una buona probabilità di urtare altri neutroni durante il rallentamento in un mantello di litio naturale. Quando un neutrone viene assorbito nel litio, esso produce trizio piú una particella alfa. Si potrebbe così ottenere un rapporto di produzione di 1,3. Se la vasca di litio avesse un raggio di almeno un metro, non vi sarebbe praticamente alcuna perdita di neutroni per fuga o assorbimento nella struttura del materiale e non vi sarebbe alcun danneggiamento del recipiente a pressione da parte dei neutroni. La scelta del litio quale fluido per l'assorbimento dell'energia dell'esplosione appare quindi particolarmente interessante.

Per quanto riguarda la scelta del materiale del recipiente a pressione, le varie prove di compatibilità indicano che i due candidati piú promettenti per essere accoppiati al litio sono il Croloy (un acciaio al cromo-molibdeno) e una lega di niobio contenente l'1 % di zirconio. La temperatura di funzionamento con il Croloy è limitata da motivi di corrosione a circa 480 °C, mentre la lega di niobio potrebbe essere usata con il litio fino ad almeno 1100 °C. Il

RECIPIENTE PER IL RECUPERO DEL DEUTERIO-TRIZIO
PREPARATORE DELLA PASTICCA
RISERVA DI COMBUSTIBILE
INIETTORE DELLA PASTICCA
POMPA
LASER
POMPA
LITIO A 430 °C
VAPORE A 450 °C
VAPORE A 210 °C
VAPORE A 450 °C
REATTORE
TURBINA AD ALTA PRESSIONE
TURBINA A BASSA PRESSIONE
GENERATORE
LITIO A 480 °C
CONDENSATORE
CALDAIA E SURRISCALDATORE DEL VAPORE
ACQUA A 38 °C
POMPA
ACQUA DI ALIMENTAZIONE A 240 °C
RISCALDATORI DI ALIMENTAZIONE
POMPA
DEAREATORE

Un impianto di potenza in formato naturale che comprende molte delle idee presentate in questo articolo potrebbe essere basato su un singolo reattore a fusione nel quale le pasticche di deuterio-trizio vengano portate all'ignizione da impulsi laser di alta potenza. L'energia delle esplosioni dovrebbe essere assorbita in un mantello di litio per la produzione del trizio e usata per produrre 150 megawatt di potenza elettrica mediante un qualsiasi ciclo a vapore di tipo completamente convenzionale.

niobio è tuttavia molto costoso e richiede una protezione dall'ossidazione; almeno nella fase iniziale di studi ed esperimenti è preferibile il Croloy.

Nella scelta del ciclo termodinamico la prima scelta logica è un ciclo Rankine a vapore. Il Croloy è largamente usato negli impianti a vapore esistenti alimentati con combustibile fossile fino a temperature del vapore di 560 °C. Cosí, se si sceglie per il sistema del litio un limite superiore di temperatura di 480 °C e se si accetta una caduta di temperatura di 28 °C tra il litio e il circuito del vapore, il reattore potrebbe essere accoppiato a un impianto a vapore del tutto convenzionale (*si veda il diagramma nella pagina a fronte*).

Si potrebbe concludere che sia necessario usare un fluido intermedio quale un sale inerte tra il litio e il circuito del vapore, a causa della grande liberazione di energia della reazione litio-acqua che potrebbe verificarsi se vi fosse una fuga da un sistema all'altro. Le conseguenze di una tale fuga non appaiono tuttavia piú gravi di quelle che si hanno negli scambiatori di calore tra il sodio-potassio e il sistema del vapore impiegati nei reattori veloci a fissione raffreddati al sodio. La quantità di radioattività nel litio dovrebbe essere piccola e la perdita di trizio in un incidente di questo tipo non dovrebbe essere preoccupante.

Per poter mantenere a un livello modesto gli sforzi di origine termica nei recipienti e nelle tubazioni del sistema del litio pare sia conveniente limitare l'aumento medio di temperatura nel litio a circa 56 °C. Ciò si potrebbe realizzare mediante una portata di circa 10 000 litri al minuto con un sistema di due pompe funzionanti in parallelo.

Una parte dell'impianto fisicamente piccola ma economicamente vitale è il sistema di recupero e di trattamento del combustibile. È stato dimostrato che tanto il trizio quanto il deuterio possono essere con poca spesa separati dal litio con un sistema di distillazione praticamente convenzionale. La miscela deuterio-trizio, dopo essere stata sottoposta a compressione, può venire liquefatta con un sistema criogenico. Si possono poi manipolare le pasticche di combustibile facendo cadere delle goccioline liquide attraverso una camera a vuoto riempita con elio gassoso freddo che dovrebbe raffreddare e congelare le goccioline. La potenza elettrica necessaria per il sistema criogenico pare sia inferiore a un watt per ogni chilowatt di potenza termica all'uscita del reattore, quindi non dovrebbe far diminuire apprezzabilmente il rendimento termico complessivo dell'impianto di potenza. È ovvio che, per man-

tenere la composizione piú opportuna delle pasticche di combustibile, sarà necessario aggiungere deuterio e asportare una miscela deuterio-trizio.

Per impianti con potenza di uscita superiore a 150 megawatt dovrebbe essere possibile rendere piú efficace il cuscino di bolle e aumentare in tal modo la potenza di uscita di un recipiente di date dimensioni di un fattore due o piú. Potenze di uscita superiori a circa 500 megawatt per recipiente richiederanno un aumento delle dimensioni e/o dello spessore del recipiente. Ciò aumenterà le difficoltà di costruzione e conseguentemente il costo unitario per chilogrammo e per chilowatt di potenza di uscita. L'alternativa è di usare un gran numero di recipienti relativamente piccoli funzionanti in parallelo (*si veda il diagramma nella pagina a fronte*). Quest'ultimo suggerimento è interessante perché renderebbe possibile la costruzione modulare con uno o piú formati standard di recipienti per reattori prodotti in serie accoppiati con pompe opportunatamente dimensionate, scambiatori di calore e tubi di collegamento. Questo metodo presenta particolari attrattive poiché le grandi turbine a vapore vengono comunemente alimentate attraverso molte tubazioni, dato che la prevista dilatazione termica consiglia di mantenere a un valore modesto il diametro – e quindi la capacità – delle singole tubazioni per vapore.

Si dovrebbe ricordare che il tipo di reattore a fusione qui proposto differisce tanto dai reattori a fissione che dagli altri tipi di reattori a fusione per il fatto che non si vede alcun vantaggio teorico nell'uso di unità molto grandi e questo è un importante punto a favore di questa idea.

Un impianto a energia di fusione innescato a laser costruito in Croloy e fatto funzionare in modo che produca vapore a 450 °C dovrebbe avere un rendimento termico globale del 40 % circa, che è all'incirca quello dei migliori impianti a combustibile fossile attualmente in funzione. Il rendimento potrebbe essere notevolmente aumentato se il sistema fosse costruito in lega al niobio e fosse fatto funzionare a una temperatura di circa 1000 °C. Tale temperatura sarebbe tuttavia troppo elevata per un ciclo a vapore poiché si avrebbe nel vapore una dissociazione dell'idrogeno e dell'ossigeno tale da provocare una seria corrosione di qualsiasi struttura metallica. Tuttavia, dovrebbe essere possibile l'impiego di un doppio ciclo a vapore fatto con un ciclo Rankine potassio-vapore che assorba calore dal litio a circa 1000 °C. Un sistema siffatto avrebbe un rendimento termico globale del ciclo del 58 % circa. Questo fatto ridurrebbe la dispersione di calore ceduto per ogni chilowatt di potenza elettrica a circa la metà di quella degli impianti a vapore convenzionali e prospetta la possibilità di rendimenti di utilizzazione termica ancora piú alti mediante integrazione con i sistemi di riscaldamento industriali e urbani.

Vi è poco interesse a sviluppare l'impianto di potenza proposto, a meno che esso non sia economicamente conveniente. Se si potesse costruire un sistema laser opportuno a un costo ragionevole – un problema piú grosso che non può essere risolto in questa fase – pare che il costo complessivo del resto del sistema non debba essere superiore al costo complessivo degli impianti piú convenzionali. Sicuramente il costo dell'impianto a vapore dovrebbe essere lo stesso di quello di impianti a vapore dello stesso tipo di uso corrente. Il circuito del litio e la caldaia del vapore dovrebbero costare all'incirca quanto l'attrezzatura corrispondente progettata per il circuito di raffreddamento al sodio dei reattori veloci fertilizzanti a fissione. Inoltre il recipiente a pressione dovrebbe essere relativamente semplice in confronto a quelli normalmente impiegati nei reattori a fissione raffreddati ad acqua. Il peso di un recipiente a pressione sferico con un diametro interno di 4 metri circa e uno spessore di pareti di 25 centimetri dovrebbe essere di circa 100 tonnellate. Nell'ipotesi di un rendimento termico del 40 % un tale impianto dovrebbe produrre una potenza elettrica di circa 60 000 chilowatt. Nell'ipotesi di un costo unitario di 3500 lire al chilogrammo, il recipiente dovrebbe costare 350 milioni, cioè circa 6000 lire al chilowatt. Il costo delle altre attrezzature, quali i sistemi criogenici per il recupero del combustibile e la preparazione delle pasticche, sembra essere una piccola frazione del costo complessivo dell'impianto.

Il costo di combustibile per il funzionamento dovrebbe essere davvero basso. Il costo del deuterio e del litio dovrebbe essere rispettivamente di circa 250 000 e 18 000 lire al chilogrammo. Ciò porterebbe a un costo di circa sole 20 lire per miliardo di joule in confronto alle attuali 400 lire per miliardo di joule degli impianti a combustibile fossile. Dal momento che la spesa di combustibili è circa la stessa di quella degli impianti convenzionali a combustibili fossili, l'impianto a energia di fusione innescato dal laser qui proposto appare economicamente conveniente anche se il costo complessivo fosse due volte maggiore.

Uno dei problemi piú importanti connessi con i reattori a fissione origina dalla grande quantità di materiale radioattivo necessariamente presente e dal conseguente pericolo potenziale che essa rappresenta. A causa di questo fatto fino al 30 % del costo complessivo di un impianto a fissione può essere assorbito per complesse tecniche atte a impedire o a limitare qualsiasi possibile incidente che possa liberare materiale radioattivo verso l'ambiente circostante. Il solo materiale radioattivo necessariamente presente nell'impianto a fusione proposto sarebbe il trizio, ma i calcoli eseguiti indicano che esso dovrebbe costituire un rischio potenziale totale assai inferiore a circa un milione di volte di quello di un reattore a fissione della stessa potenza. Ciò dovrebbe ridurre i costi e facilitare notevolmente i problemi di installazione.

Lo studio dei progetti di impianti a energia di fusione innescati a laser fanno apparire attraente l'idea e sollevano perciò degli interrogativi riguardo ai problemi di sviluppo. I piú importanti sono ovviamente relativi al laser e al processo di ignizione della pasticca. Si può costruire un laser sufficientemente potente e utilizzarne gli impulsi per l'ignizione della pasticca? E se cosí è, la resa di energia di fusione sarà molte volte superiore all'energia all'ingresso del laser? Con quale rendimento si può attenuare l'onda d'urto mediante l'immissione di una grossa frazione di bolle di gas in una vasca di litio in rapida rotazione? Questi sono i principali interrogativi ai quali si deve dare sperimentalmente una risposta prima di poter dire se l'idea proposta è davvero realizzabile in pratica. Vi sono ovviamente molti altri problemi di sviluppo, ma nessuno di essi sembra cosí difficile da sollevare dubbi sulla realizzabilità dell'idea.

Un impianto di potenza da un migliaio di megawatt potrebbe essere composto da una serie di sedici moduli costituiti da un reattore a fusione con il proprio scambiatore di calore e il proprio sistema di pompe. Il sistema potrebbe produrre vapore a una temperatura di 450 °C e a una pressione di circa 80 atmosfere in grado di far muovere una combinazione di turbine a pressione alta e intermedia e con un ulteriore riscaldamento a 450 °C per una coppia di turbine a bassa pressione. Con l'aggiunta di un ciclo a vapore di potassio operante all'ingresso della turbina a una temperatura di 1000 °C l'efficienza termica complessiva dell'intero sistema potrebbe raggiungere il 58 %.

Le bolle magnetiche

Ogni materiale magnetico è suddiviso in zone con diverse direzioni di magnetizzazione. Questi domini magnetici assumono la forma di piccole bolle che vengono impiegate in un nuovo tipo di memoria per calcolatori

di Andrew Bobeck e H.E.D. Scovil

Le prestazioni dei calcolatori elettronici e in generale di tutti gli elaboratori di dati dipendono dalla velocità, dalla capacità e dalla regolarità dei loro sistemi di memoria. A loro volta, queste caratteristiche dipendono essenzialmente dai costi che si è disposti ad affrontare. I sistemi di memoria piú veloci e piú flessibili sono quelli che utilizzano dei minuscoli nuclei anulari di ferrite congiunti con una maglia di sottili fili metallici oppure, in epoca piú recente, quelli che impiegano dei circuiti a transistor costruiti su minuscole piastrine (*chips*) di silicio. Attualmente i nuclei sono relativamente poco cari: circa cinquanta-cento lire per bit (cifra binaria) di capacità di immagazzinamento. Anche un costo cosí basso è comunque notevole: una memoria capace di raddoppiare le capacità del cervello umano – stimate a circa 10^{10} bit – avrebbe un costo di circa 10^{11} lire. Una centrale telefonica elettronica automatica ha un fabbisogno piú modesto, circa 10^{7} bit, e il prezzo della memoria ammonterebbe a 10^{8} lire. Se si pensa che, per assicurare un funzionamento regolare, una centrale telefonica deve disporre di una capacità doppia, ci si rende conto dell'importanza che può avere per le compagnie telefoniche (e non solo per esse) la scoperta di sistemi elettronici automatici meno costosi.

È attualmente allo studio, presso i Laboratori della Bell Telephone, una interessante alternativa che prevede la messa a punto di una nuova tecnologia per cui i bit vengono immagazzinati sotto forma di « bolle » magnetiche in movimento entro sottili pellicole di materiale magnetico. In realtà, queste bolle non sono altro che dei domini magnetici cilindrici che hanno una polarizzazione opposta a quella della pellicola magnetica entro la quale si muovono. Queste bolle sono stabili rispetto a tutta una serie di condizioni fisiche e possono muoversi a velocità assai elevate.

I dati attualmente a nostra disposizione ci permettono di affermare che le memorie a bolla magnetica saranno molto piú economiche di quelle a nucleo e sino a dieci volte piú veloci degli attuali sistemi a disco magnetico largamente impiegati negli elaboratori ad alta capacità di memoria. Il funzionamento di questi sistemi si basa sul movimento meccanico di un disco sotto una testina capace di leggere dei dati precedentemente registrati o di incidere dei nuovi dati. Nei migliori sistemi a disco magnetico, che hanno piú tracce per ogni disco, i vari dischi girano in continuazione sotto le testine che sono in numero uguale al numero delle tracce. Pertanto i sistemi a disco permettono di accedere a ogni serie di dati registrati nel limite di tempo necessario perché le informazioni vengano passate sotto le testine. Un inconveniente di questo sistema è che non è possibile manipolare un'informazione già registrata senza prima averla letta e reincisa nuovamente: questo procedimento richiede un certo tempo. Comunque, dato che i sistemi a disco sono meccanici, non hanno la stessa affidabilità dei sistemi completamente elettronici che vengono installati nei centralini telefonici e che sono progettati per funzionare 40 anni.

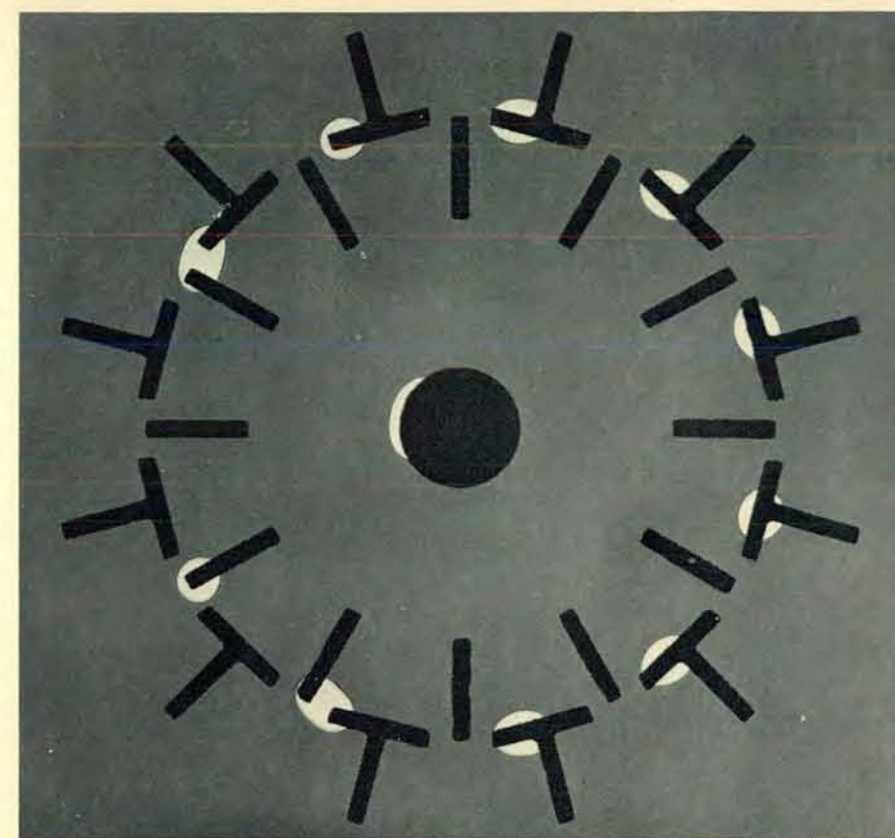
Le memorie a bolla magnetica, attualmente in fase di sviluppo, sono per molti aspetti l'analogo completamente elettronico delle memorie a disco magnetico, anche se esistono notevoli differenze tra i due sistemi. In entrambi i sistemi, l'informazione viene immagazzinata sotto forma di stato di magnetizzazione su (o in) una sottile pellicola magnetica. Però, mentre nelle memorie a disco la pellicola viene mossa meccanicamente a grande velocità, nelle memorie a bolla le bolle stesse si spostano ad alta velocità attraverso la pellicola; nel primo sistema, inoltre, l'informazione è rigidamente fissata e si può muovere sotto la testina di lettura-scrittura soltanto in una dimensione mentre nel secondo sistema essa può spostarsi in ogni punto della pellicola, cioè in due dimensioni. Con il sistema a bolla è anche possibile eseguire parecchie operazioni logiche sui dati immagazzinati senza doverli rileggere e riscrivere nuovamente. Infine, poiché i sistemi a bolla non contengono parti mobili, essi possono, almeno in linea di principio, funzionare per molti anni, anche se ovviamente ciò è ancora da dimostrare.

Per lo sviluppo dei sistemi a bolla magnetica è stato necessario attingere a tutto il patrimonio della teoria del magnetismo sviluppata da cinquant'anni a questa parte e alle tecniche piú recenti sulla produzione di materiali magnetici particolari. Uno dei concetti chiave della teoria consiste nell'idea di dominio magnetico, presagita già da Maxwell, secondo cui « ogni magnete è costituito da un insieme di particelle ciascuna delle quali è a sua volta un piccolo magnete ». Come altri fisici del XIX secolo, egli tentò di spiegare le proprietà macroscopiche dei materiali ferromagnetici sulla base delle interazioni elettromagnetiche elementari tra queste « particelle » magnetiche di dimensioni molecolari. Nonostante l'acutezza del modello, esso tuttavia fallí quando si trattò di interpretare alcune proprietà fondamentali dei magneti.

Nel 1907 Pierre Ernst Weiss comprese che alcune incongruenze della teoria potevano essere superate facendo l'ipotesi che, all'interno di ogni materiale ferromagnetico, esistano degli intensi « campi molecolari » capaci di provocare un allineamento dei magneti molecolari. Weiss calcolò che l'intensità del campo necessaria a produrre questo allineamento era di molto superiore all'intensità che si poteva ottenere con gli ordinari elettromagneti in laboratorio: essa doveva essere all'incirca di 10 milioni di oersted ma lo stesso Weiss non riusciva a spiegare come potessero essere prodotti campi tanto intensi. Vent'anni piú tardi, la meccanica quantistica permise di concludere che i campi molecolari ipotizzati da Weiss avevano origine dalle forze di scambio tra gli atomi ferromagnetici e che i momenti molecolari elementari non erano altro che i momenti magnetici atomici associati agli spin degli elettroni.

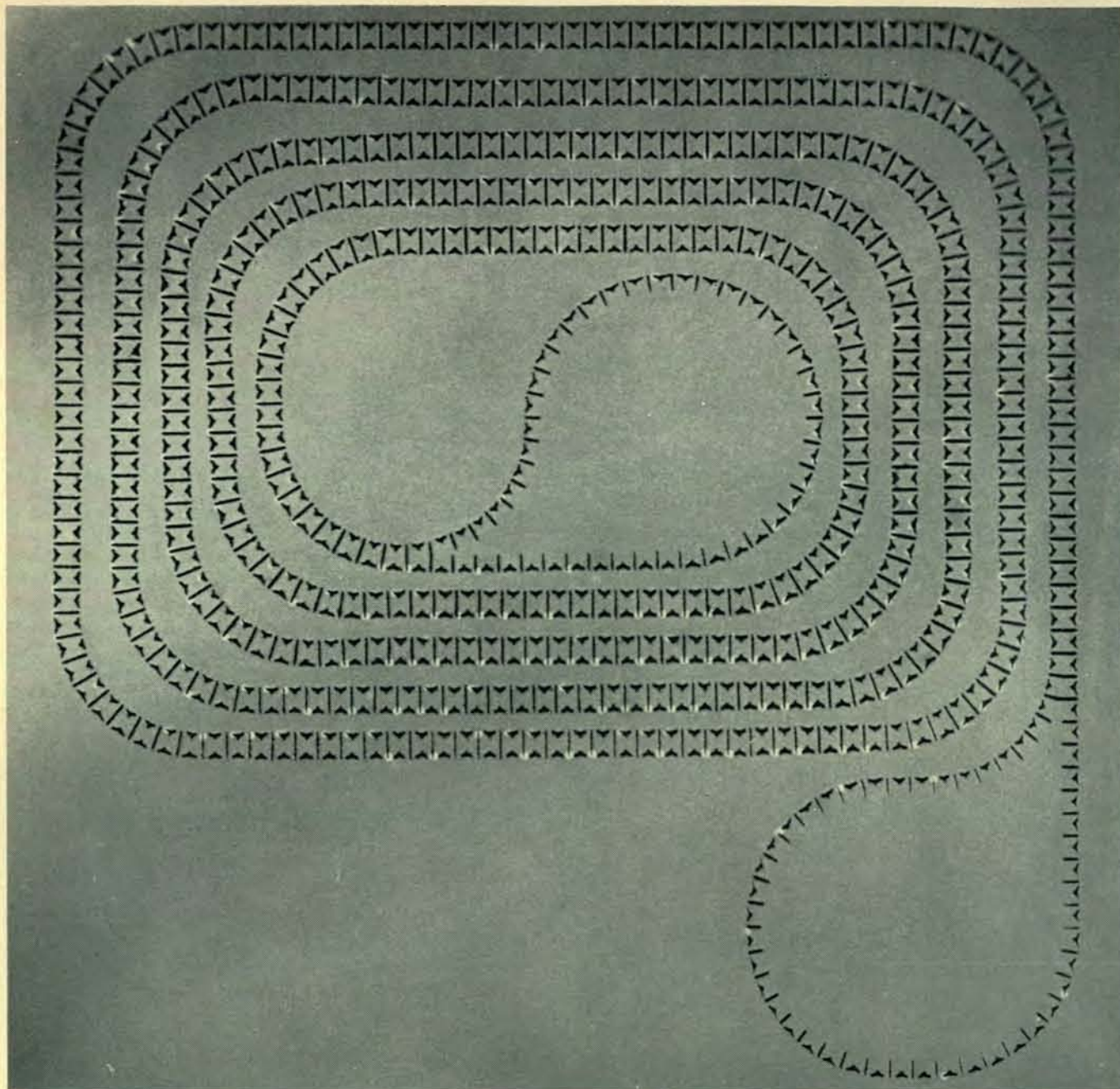
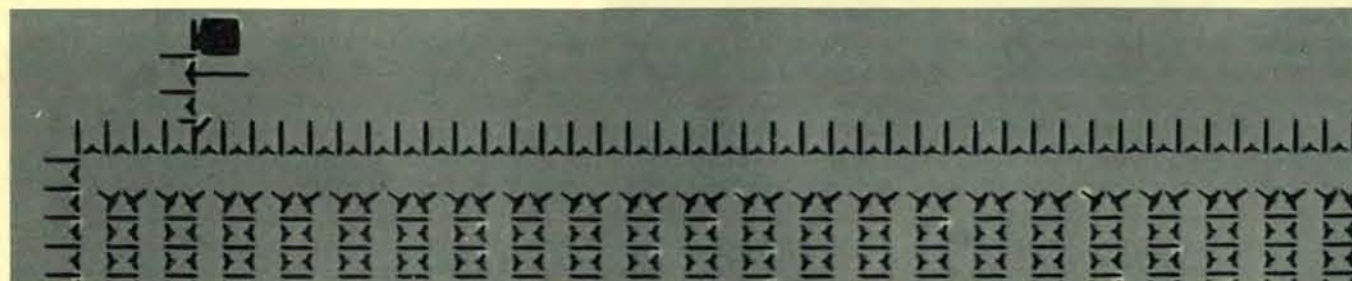
Anche senza queste cognizioni Weiss intuí che non era indispensabile che i magneti molecolari si allineassero in tutto il volume del materiale magnetico; egli suggerí che soltanto delle piccole porzioni, i « domini magnetici », fossero spontaneamente magnetizzate. A prima vista, sembrò plausibile che in un materiale costituito da molti cristalli elementari, questi domini potessero essere cosituiti dai singoli cristalli, ma questa ipotesi fu scartata quando fu possibile osservare i domini magnetici anche in singoli cristalli. Ora sappiamo che la forma e le dimensioni dei domini sono essenzialmente determinate dall'equilibrio di piú forze che rendono minima la somma dell'energia magnetica, dell'energia di scambio e dell'energia magnetocristallina (*si vedano le figure alle pagine 98 e 99*).

Poiché i domini magnetici costituiscono la base della nuova tecnologia delle bolle magnetiche, descriveremo piú dettagliatamente questo processo di minimizzazione. L'intensa forza di scambio che è capace di allineare i momenti magnetici è essenzialmente isotropa, vale a dire, non ha direzioni privilegiate. I cristalli dei materiali magnetici hanno, invece, dei campi interni che provocano l'allineamento dei momenti magnetici secondo direzioni privilegiate, generalmente quelle corrispondenti agli assi del reticolo cristallino. Se il cristallo è soggetto a un campo esterno la cui direzione non coincide con una di queste direzioni preferenziali, il momento magnetico tende ad allinearsi secondo una direzione intermedia che dipende dalla risultante delle forze interne ed esterne.

Esiste inoltre anche una terza forza in gioco che deriva dalle discontinuità nella magnetizzazione che producono dei campi magnetostatici. I campi magnetostatici agiscono sempre in senso contrario alla magnetizzazione e quindi, oltre a ridurre la magnetizzazione del ferromagnete, riducono anche la loro stessa intensità: per questo motivo essi vengono chiamati campi autosmagnetizzanti. Questi campi sono sempre presenti anche nei cristalli apparentemente perfetti e manifestano la loro presenza attraverso la formazione di domini magnetici.

È possibile stabilire un confronto tra i domini magnetici lineari e le bolle magnetiche esaminando questa microfotografia e quella raffigurata in copertina. La configurazione a *T* e a barre rappresenta una variante dello schema delle due figure della pagina seguente. La figura della copertina mostra che, in assenza di un campo magnetico esterno, i domini di opposta magnetizzazione occupano delle superfici di forma articolata e di uguale estensione. Si possono osservare, anche se con una certa difficoltà, le sottili pareti di separazione tra i domini. In questa figura la presenza di un campo magnetico ha provocato una contrazione dei domini sino a ridurli in bolle. Esse si muovono nel piano della figura per mezzo di un campo magnetico rotante.

Tagliando un cristallo ferromagnetico secondo una sezione perpendicolare al suo asse preferenziale di magnetizzazione e osservandolo attraverso un polarizzatore, si può facilmente verificare la presenza di domini magnetici che si manifestano sotto forma di strisce ondulate. A seconda della loro polarità magnetica, i domini fanno ruotare il piano di polarizzazione della luce polarizzata in direzione opposta. Agendo sul filtro polarizzatore del microscopio, si può fare in modo che una metà dei domini risulti chiara e l'altra metà scura.

I limiti di separazione tra domini adiacenti si dicono pareti del dominio: anche questi hanno origine da un processo di minimizzazione dell'energia. Il contorno di ogni dominio non è rappresentato da una linea ben definita ma si estende attraverso alcune centinaia di piani atomici del cristallo (*si veda l'illustrazione a destra a pagina 99*). Questo fatto è il risultato di un riassestamento che rende minimo il lavoro necessario per ruotare i momenti atomici contro le forze di scambio (che tendono a tenere paralleli i momenti adiacenti) e cosí pure il lavoro necessario per far ruotare i momenti rispetto alla direzione dell'asse di facile magnetizzazione del campo magnetocristallino. Quanto piú intensa è la forza di scambio, tanto piú spesso è il contorno del dominio; lo spessore diminuisce invece all'aumentare dell'anisotropia magnetocristallina.

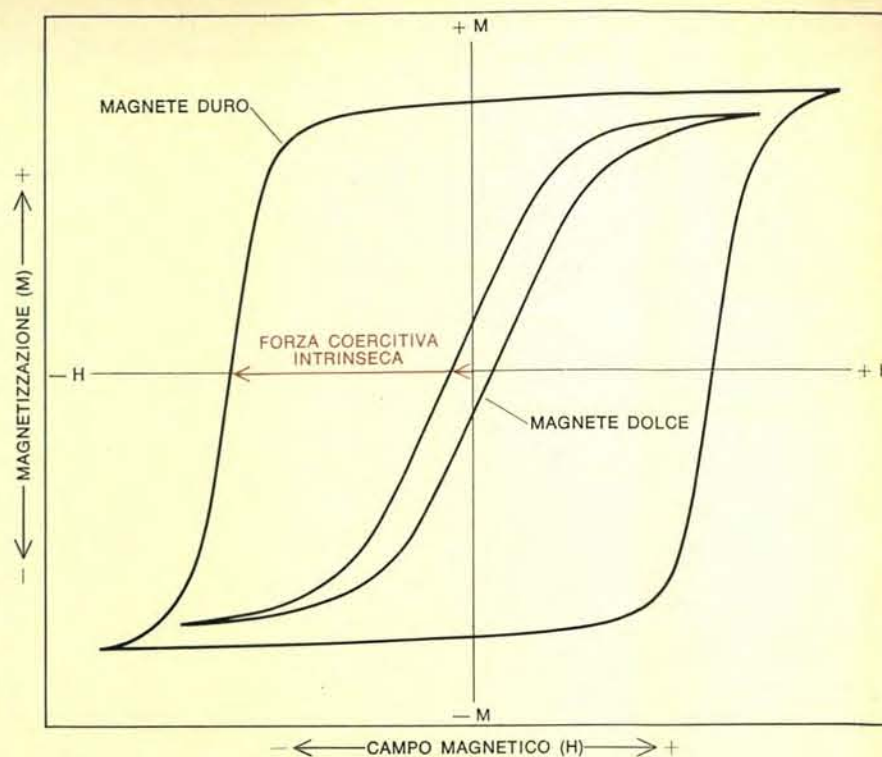
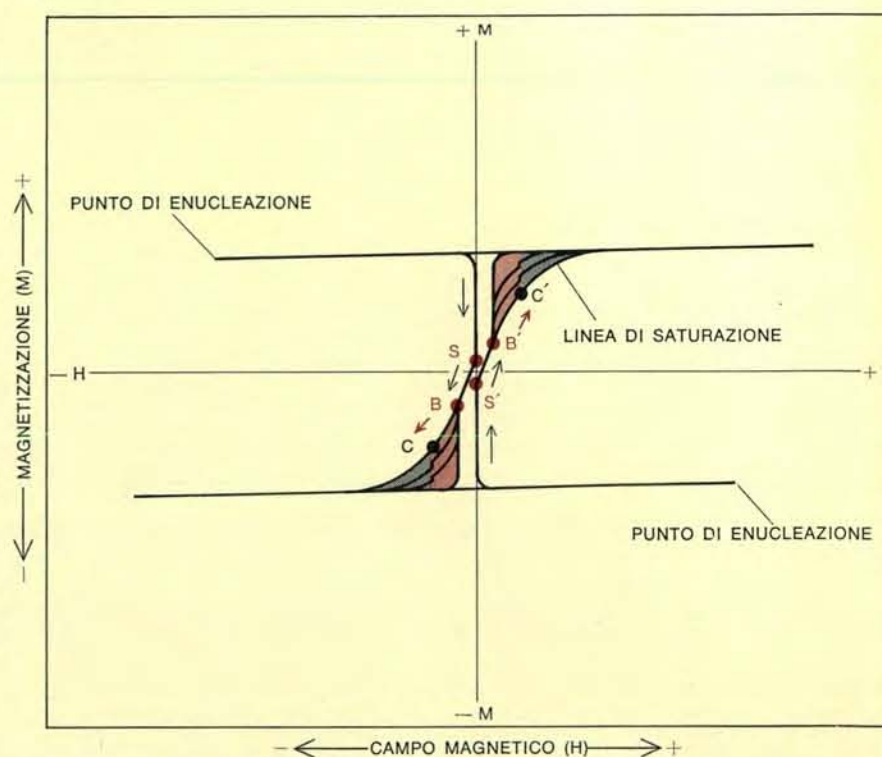
Si possono formare domini di forma geometrica semplice soltanto in cristalli singoli contenenti pochi difetti: i materiali ordinari contengono generalmente piccoli difetti, inclusioni, vacanze e tensioni che producono dei campi magnetostatici altamente localizzati e delle discontinuità che agiscono sui contorni dei domini impedendo loro di assumere la configurazione di minima energia. Questi difetti rappresentano comunque un vantaggio nel caso in cui si

Nei Laboratori della Bell Telephone è stata costruita questa memoria a bolla il cui schema assomiglia alle rotaie di un trenino elettrico in miniatura. Le bolle magnetiche vengono trasportate lungo le «rotaie» per mezzo di elementi di Permalloy (magneti dolci) posti sulla superficie di una lamina magnetica sotto forma di *Y* tronche e di sottili barre. Ogni coppia *Y*-barra costituisce una fase della memoria: in complesso il numero delle fasi è di 1074. La circolazione delle bolle avviene per mezzo di un campo magnetico rotante posto nel piano della memoria. La direzione di circolazione viene invertita nella parte centrale e nella parte inferiore destra. Ogni bolla ha la capacità di rappresentare una cifra binaria di informazione.

In una piastrina di due millimetri di lato è contenuta una memoria a diecimila fasi. In questa microfotografia è rappresentato soltanto il 7 % circa dell'intero complesso. Come nella piccola memoria della figura in alto, le bolle vengono guidate per mezzo di un sistema di *Y* e di sottili barre di Permalloy. Le bolle sono appena visibili e sono rappresentate dai punti chiari racchiusi nella regione compresa tra i due tipi di struttura. Il quadrato scuro nella parte superiore sinistra della microfotografia rappresenta il generatore di bolle che fornisce appunto le bolle necessarie alla memoria. Un rivelatore, non rappresentato nella figura, assolve la funzione di leggere a comando i dati binari che sono stati introdotti nel flusso di bolle.

vogliano ottenere dei materiali (come per esempio i magneti permanenti) per i quali si richiedono delle alte forze coercitive e delle grandi isteresi.

Per gli usi di cui stiamo trattando invece, in cui è necessaria la massima mobilità dei domini, la coercitività del materiale, o « durezza » magnetica, deve essere bassa e facilmente controllabile (*si vedano le illustrazioni in questa pagina*).

Consideriamo ora un sottile strato di un materiale magnetico anisotropo ottenuto tagliando un singolo cristallo sintetizzato: se si osserva questo strato in assenza di campi magnetici esterni per mezzo di luce polarizzata, si vede una figura a strisce ondulate che rappresentano i domini. In una metà delle strisce i minuscoli magneti interni sono orientati verso l'alto, mentre nell'altra metà sono diretti verso il basso. Inoltre, secondo come è orientato il filtro polarizzatore, una parte delle strisce appare chiara mentre l'altra è scura: i due sistemi di strisce occupano la stessa porzione di superficie (*si vedano le illustrazioni a pagina 106*).

Se ora si sottopone lo strato a un campo magnetico a esso perpendicolare e si aumenta progressivamente l'intensità del campo, si può osservare che le strisce ondulate la cui magnetizzazione è opposta a quella del campo tendono a ravvicinarsi. Questo processo continua sino a che tutti i domini « isolati » (e cioè quelli che non sono vincolati ai bordi dello strato) non subiscono un'improvvisa contrazione assumendo la forma di piccoli cerchi che prendono appunto il nome di bolle: viste di fronte, queste bolle sono in realtà dei cilindri. Se l'intensità del campo esterno viene ulteriormente aumentata cominciano a restringersi, sino a scomparire del tutto.

Ogni bolla si comporta come un piccolo magnete fluttuante in un campo magnetico di opposta polarità. È possibile verificare la grande mobilità delle bolle muovendo un sottile filo magnetizzato lungo la superficie dello strato e osservando al microscopio che effettivamente è possibile spostare le bolle in ogni direzione. D'altra parte, poiché hanno tutte la stessa polarizzazione si respingono mantenendo tra loro una distanza praticamente uniforme. Si può dimostrare che, se il campo esterno che produce le bolle varia, sia positivamente che negativamente, entro intervalli non superiori al 20 %, le bolle sono completamente stabili e posseggono una assoluta mobilità. Abbiamo quindi realizzato su scala microscopica degli oggetti che sono altrettanto durevoli e impenetrabili delle palle da biliardo ma che hanno in piú il vantaggio di respin-

I materiali magnetici hanno un meccanismo di risposta ai campi magnetici magnetizzanti o smagnetizzanti che dipende dalla loro maggiore o minore « durezza ». Queste curve di isteresi rappresentano i due diversi tipi di comportamento: i magneti duri, a elevata coercitività, sono insensibili alle variazioni di direzione della magnetizzazione mentre i magneti dolci vengono facilmente magnetizzati o smagnetizzati.

I materiali che vengono utilizzati per produrre le bolle sono caratterizzati da una debole magnetizzazione (*M*) e da una bassa coercitività. Se si produce una inversione della direzione di magnetizzazione anche in una piccola regione della lamina in entrambi i punti di nucleazione, l'intera lamina si smagnetizza rapidamente. La lamina neutra verrà pertanto divisa in due domini lineari in corrispondenza del punto *S* (o *S'*). Se viene applicato un debole campo magnetico esterno $-H$, in corrispondenza del punto *B* cominceranno a formarsi delle bolle che rimangono stabili sino a che il campo non abbia raggiunto il valore *C*. Quando lo strato smagnetizzato viene sottoposto a un campo magnetico opposto ($+H$) il tragitto è quello diretto verso la parte superiore, vale a dire dal punto *S'* sino al punto *B'* e quindi al punto *C'*.

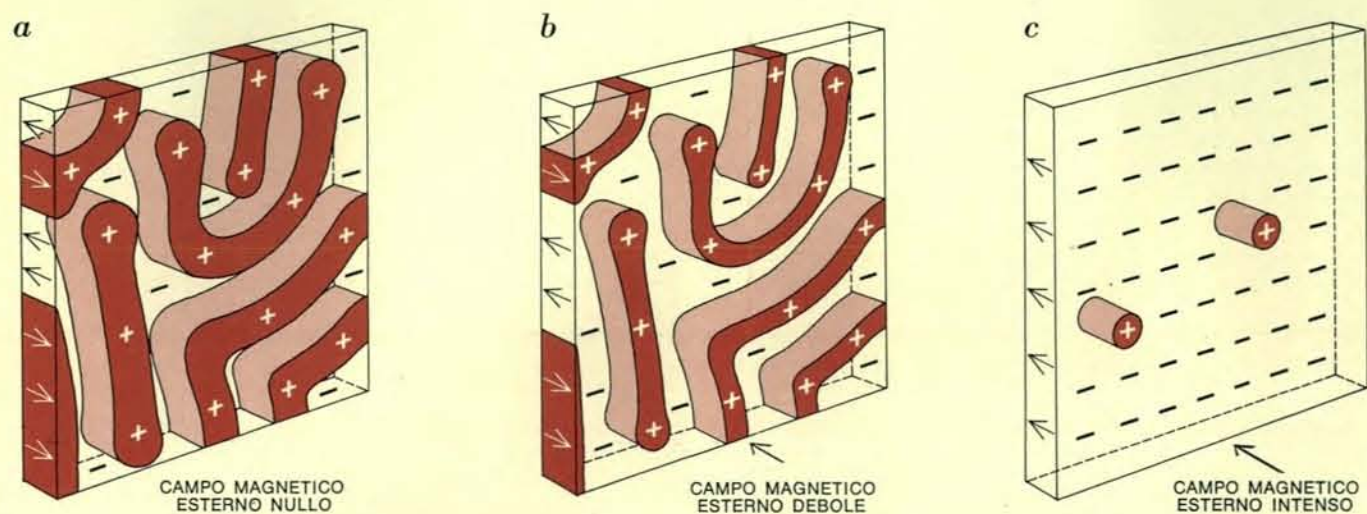
gersi l'un l'altro. Inoltre le bolle possono essere create ovunque ve ne sia la necessità e distrutte per mezzo delle tecniche che descriveremo piú avanti.

I primi materiali magnetici che avevano le proprietà adatte allo studio delle nuove tecnologie a bolla risultarono essere le ortoferriti, e cioè una particolare classe di ferriti dalla formula chimica $RFeO_3$ in cui R rappresenta l'ittrio o un altro degli elementi delle terre rare. L'ortoferrite a samario-terbio ne è un esempio. È possibile ottenere le ortoferriti sotto forma di singoli cristalli mescolando opportunamente gli ingredienti e facendo fondere la miscela in un crogiuolo, lasciandola poi raffreddare per un periodo di alcune settimane. Generalmente, terminato questo procedimento, si ricavano alcuni cristalli della grandezza desiderata.

Nel tentativo di ricavare dei materiali che potessero dar luogo a piccole bolle dotate di grande velocità, abbiamo esaminato, durante gli ultimi anni, centinaia di lamine di ortoferriti di diversa composizione. L'obbiettivo minimo che ci siamo prefissi è stato quello di ottenere una densità dell'ordine del milione di bolle al centimetro quadrato e una mobilità che ci permettesse di disporre di un'efficienza di un milione di bit al secondo. Il migliore dei materiali considerati (l'ortoferrite a samario-terbio) soddisfa al secondo di questi requisiti ma raggiunge densità dieci volte inferiori a quelle necessarie perché le sue bolle hanno delle dimensioni troppo grandi: circa 25 micron di diametro. (La distanza ottimale tra le bolle è di quattro diametri da centro a centro). Le piú piccole bolle finora osservate vengono prodotte in una ferrite con struttura cristallina esagonale ($PbFe_{12}O_{19}$); le sue bolle hanno un diametro di un micron e realizzano quindi delle densità di circa 10 milioni al centimetro quadrato ma purtroppo la loro velocità è inferiore di un fattore 20 a quella desiderata.

Dopo piú di un anno di ricerche trovammo il materiale che faceva al caso nostro: dei granati sintetici dalla formula generale $A_3Fe_5O_{12}$ in cui il simbolo A rappresenta l'ittrio, qualunque altra terra rara, oppure, almeno in parte, lantanio o bismuto (*si veda l'illustrazione in basso a pagina 103*). I granati sono facilmente trattabili chimicamente e i parametri caratteristici delle bolle sono facilmente controllabili. Nei granati piú perfetti, il diametro delle bolle è di circa tre micron il che permette di ottenere densità dell'ordine del milione al centimetro quadrato. In questi granati, le bolle possono spostarsi a una velocità di almeno un milione di intervalli al secondo. Inoltre, l'energia necessaria per spostare nel granato una bolla di quattro diametri è di circa 4×10^{-14} joule, vale a dire circa un duecentocinquantesimo dell'energia necessaria per far scattare il piú piccolo transistor. Quindi, per spostare un milione di bolle un milione di volte al secondo, il che equivale a 10^{12} operazioni, è necessaria una potenza di soli 40 milliwatt contro una potenza di 10 watt necessaria per ottenere la stessa efficienza con i transistor.

L'idea di utilizzare delle particelle magnetiche per accumulare dati e per eseguire operazioni logiche può apparire abbastanza ovvia oggi che sono note le proprietà delle bolle magnetiche, ma in effetti bisogna considerare che i primi tentativi di utilizzare i domini magnetici a questo scopo seguirono delle vie completamente diverse. La prima idea fu quella di ottenere delle minuscole regioni che avessero una magnetizzazione invertita entro un filamento che avesse gli assi di magnetizzazione preferenziale paralleli alla propria lunghezza. In questo modo ognuna di queste regioni sarebbe stata isolata per tut-

L'effetto dei contorni dei domini è quello di ridurre la quantità di energia magnetostatica in un cristallo magnetico. Un campo magnetico circonda ogni piccolo magnete elementare come per esempio accade nel « monopolo » magnetico positivo in *a*. Se accanto a questo viene posto un altro monopolo avente polarità opposta (*b*), i due campi tendono a elidersi e l'energia magnetostatica diminuisce. Se un singolo cristallo magnetico avente la forma di una sbarra (*c*) è uniformemente magnetizzato, i poli opposti sono relativamente distanti e, pertanto, l'energia magnetostatica ha un valore massimo. Se a metà della sbarra la magnetizzazione venisse invertita (*d*), tramite la divisione della sbarra stessa in due domini, l'energia magnetostatica verrebbe considerevolmente ridotta. La parete del dominio, ossia la superficie di separazione tra le due zone (*in grigio*), introduce comunque una certa quantità di energia. Il sistema tende verso uno stato di energia minima in cui la diminuzione della energia magnetostatica è compensata dall'aumento della energia della parete del dominio. Nella curva qui sopra, l'energia minima corrisponde a una o a due suddivisioni.

La stabilità delle bolle magnetiche è dovuta all'applicazione di un campo magnetico perpendicolare alla lamina magnetica che, in caso contrario, assumerebbe la configurazione di minima energia in cui i domini assumono la forma a serpentina (*a*). La lamina viene tagliata in modo tale che l'asse preferenziale di magnetizzazione sia perpendicolare alla superficie. Il campo magnetico esterno produce una variazione dell'energia totale della lamina in modo tale che i domini aventi polarità opposta a quella del campo esterno si contraggono (*b*) sino ad assumere una configurazione cilindrica (*c*): questi cilindri vengono chiamati bolle magnetiche. La curva di destra rappresenta i contributi dovuti ai tre tipi di energie (quella dovuta alle pareti dei domini, quella del campo esterno e l'energia magnetostatica); la somma di questi contributi dà luogo a una curva che rappresenta l'energia totale e che ha approssimativamente la forma di una scodella. Il raggio di stabilità della bolla corrisponde al valore minimo dell'energia totale, al « fondo » della scodella. Quando il campo magnetico esterno raggiunge un valore critico, le bolle si disintegrano.

I contorni dei domini sono delle zone di transizione presenti in tutti i materiali magnetici attraverso le quali la magnetizzazione inverte la sua direzione. L'effetto combinato della forza esercitata dai contorni del dominio e degli effetti del campo magnetico esterno (*frecce a destra*), tende a ridurre il diametro delle bolle.

Uno degli schemi di trasporto delle bolle prevede che sopra lo strato magnetico venga posto un sistema di spire conduttrici. Il verso della corrente elettrica viene scelto in modo di far diminuire il valore del campo magnetico all'interno della spira e quindi di provocare una forza attrattiva direttamente sulle bolle. Quando le spire vengono alimentate in successione la bolla si sposta da una spira verso la successiva.

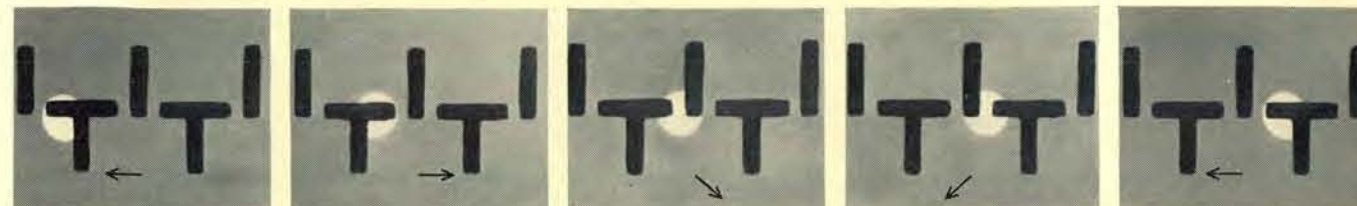
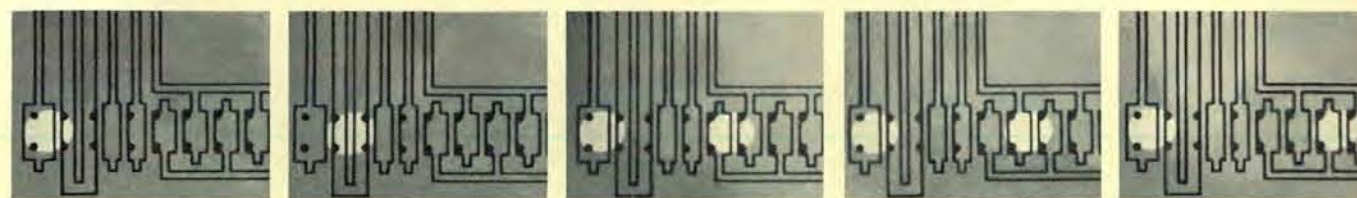
ta la sua estensione dal contorno di un dominio. Avvolgendo il filamento entro bobine di guida, sarebbe stato possibile spostare le regioni magnetiche ad alta velocità, come il petrolio pompato in un oleodotto. Nel codice binario la presenza di una regione rappresenta 1, mentre la sua assenza rappresenta 0. Apposite testine di scrittura possono introdurre i dati nel circuito mediante la creazione di zone magnetiche aventi distanze prefissate; delle testine di lettura poste a valle possono estrarre i dati mentre questi passano a grande velocità. In laboratorio è stato dimostrato che si possono ottenere delle efficienze di 600 000 bit al secondo ma è presumibile che si possano ottenere efficienze anche di un milione di bit.

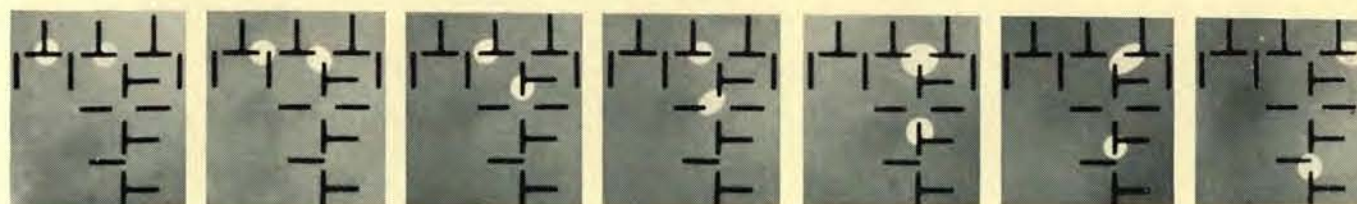
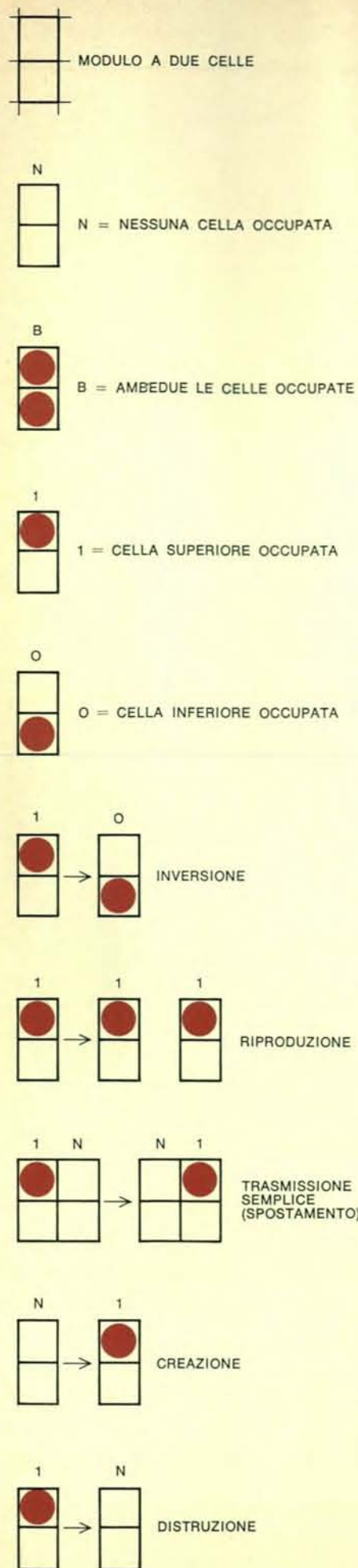
Questi condotti magnetici posseggono molte interessanti proprietà. Lo sviluppo delle ricerche dimostrò che essi potevano essere prodotti in modo abbastanza semplice e con costi relativamente bassi. Un grave inconveniente di questo sistema era rappresentato dal fatto che non era possibile far muovere le zone tra filamenti diversi se non leggendone una in un filamento e poi registrandola nuovamente in un altro. Di conseguenza, molti sistemi di immagazzinamento che a prima vista sembravano abbastanza semplici, nella pratica si dimostrarono macchinosi e costosi. Un inconveniente ancora piú grave era rappresentato dalle dimensioni: i contorni dei domini a ogni confine di zona magnetica non erano ad angolo retto rispetto al filamento ma si estendevano davanti e dietro a forma di coni. La lunghezza totale di una zona è circa cento volte il diametro del filamento; con filamenti molto sottili è possibile raggiungere una densità che si aggira sui 10 bit al centimetro, ma la densità che si raggiunge è di circa la metà.

Nell'affrontare questi problemi insieme ai nostri collaboratori U.F. Gianola, William F. Shockley, P.C. Michaelis e R.C. Sherwood ci chiedemmo se sarebbe stato possibile creare dei domini compatti all'interno di lamine di materiale magnetico, anziché all'interno di un filamento magnetico, e di farli muovere nelle due dimensioni. A questo punto alcuni di noi si resero conto della incredibile varietà di configurazioni che potevano assumere i domini magnetici, peraltro perfettamente note a coloro che si occupavano normalmente di materiali magnetici. Anche se in effetti avevamo già avuto modo di vedere occasionalmente al microscopio quelle che in seguito avremmo definito col nome di bolle, non ci venne in mente che queste bolle avrebbero potuto essere utilizzate come portatrici di dati. Furono necessari molti mesi, nei quali cercammo di elaborare dei procedimenti per creare i portatori magnetici all'interno delle lamine di materiali magnetici, prima di accorgerci che le bolle si celavano nei nostri campioni pronte per essere utilizzate. La nostra eccitazione crebbe alquanto quando ci accorgemmo che queste particelle estremamente mobili erano abbastanza stabili, che esercitavano tra loro delle forze repulsive e che si potevano far viaggiare liberamente all'interno del materiale esercitando delle forze sui loro contorni di dominio.

Descriveremo ora alcune delle tecniche con le quali è possibile utilizzare le bolle per accumulare dati e per eseguire operazioni logiche. Il controllo delle bolle richiede la creazione di campi magnetici guida, ossia di campi che abbiano una componente lungo il pia-

Per spostare le bolle ad angolo retto in ogni punto dello strato si può utilizzare questa rete di spire conduttrici. La struttura dell'impianto è tale per cui tutte le spire collegate con il circuito X_1 vengono alimentate contemporaneamente; lo stesso accade successivamente per le spire collegate al circuito X_2 e al circuito X_3. Allo stesso modo, vengono alimentati in successione i circuiti di spire Y_1, Y_2 e Y_3. Se inizialmente una bolla magnetica è nella posizione A, la successione X_1, X_2, X_3 provoca lo spostamento della bolla nella posizione *B* mentre la successione Y_1, Y_2, Y_3 la sposterà da *B* a *C*.

Questo sistema di trasporto delle bolle viene denominato «metodo di propagazione con guida a campo» in quanto le bolle si spostano a causa della variazione del campo magnetico esterno piuttosto che a causa della corrente elettrica che attraversa dei conduttori. Un sistema di struttura di sottili pellicole di Permalloy viene posta sotto lo strato contenente le bolle: queste pellicole formano una serie di teste di frecce. Allo scopo di provocare alternativamente la contrazione e l'espansione delle bolle, il campo magnetico esterno viene ritmicamente aumentato e diminuito. Ogni volta che una bolla si contrae, si sposta verso la punta della freccia corrispondente, mentre a ogni espansione essa si sposta in corrispondenza della testa di freccia successiva.

Un altro schema di guida a campo è rappresentato da questo trasportatore di bolle avente una struttura a *T* e a barre. Come indicano le frecce nere, il campo magnetico esterno ruota di un angolo di 360 gradi in senso orario nel piano dello strato. Quando il campo compie le rotazioni, le diverse regioni delle *T* e delle barre, costruite in Permalloy, vengono successivamente polarizzate in senso negativo e in senso positivo. Nell'ipotesi in cui la parte superiore della bolla sia polarizzata negativamente, essa verrà attratta verso i poli positivi in misura che questi si spostano verso destra attraverso la struttura a *T* e barre. Con questo metodo si è riusciti a ottenere degli spostamenti in campi rotanti a una velocità di un milione di cicli al secondo.

Questo partitore di bolle è costituito da una particolare struttura contenente un circuito a forcina che può venire alimentato separatamente quando la bolla si trova sopra di esso (*secondo fotogramma a sinistra*). Quando la corrente circola nel circuito a forcina la bolla viene separata in due a una velocità talmente grande che non è possibile fotografarla (il tempo impiegato è di pochi milionesimi di secondo). In seguito a ciò una bolla della nuova coppia si sposta immediatamente verso destra.

Il generatore di bolle è un apparato di guida a campo capace di creare una nuova bolla tutte le volte che un campo complanare a esso compie una rotazione. Il dominio che appare come una mezzaluna sporgente sotto il disco nero indica la direzione del campo. La bolla, che ha inizialmente la forma di un filamento, assume poi la forma circolare quando avviene la separazione dal generatore. In seguito, essa viene trasportata verso destra per mezzo di una configurazione *Y*-barre modificata.

Questo circuito separatore contiene un sistema capace di catturare una bolla ogni due. Qui due bolle si avvicinano al sistema da sinistra e la prima viene costretta nell'ingranaggio. La presenza di questa bolla nell'ingranaggio costringe la seconda a rimanere nella pista orizzontale. A sua volta, il passaggio della seconda bolla costringe la prima a uscire dall'ingranaggio.

no della lamina. I metodi che si possono usare sono essenzialmente due. Il primo consiste nell'impiego di conduttori percorsi da una corrente che genera il campo magnetico voluto ed è chiamato metodo di guida a conduttore. Il secondo sistema, di guida a campo, consiste nell'immergere l'intero strato in un campo magnetico pulsante o rotante che agisce sulle bolle per mezzo di opportuni materiali magnetici localizzati in modo da concentrare gli effetti del campo.

Il metodo di guida a conduttori implica che (per mezzo di tecniche fotografiche) vengano disposti sulla lamina dei sistemi di conduttori contenenti delle spire che abbiano approssimativamente lo stesso diametro delle bolle. L'intensità del campo magnetico generato da ogni spira è abbastanza elevata in modo che la bolla piú vicina possa venire attratta verso il suo centro (*si veda l'illustrazione in alto a pagina 100*). Come si può osservare, se le spire adiacenti vengono successivamente eccitate, una bolla può passare agevolmente da una spira alla successiva; inoltre, affinché una bolla possa spostarsi nella direzione desiderata, le spire vengono eccitate da tre circuiti guida indipendenti ciascuno dei quali viene collegato con una spira ogni quattro spire della serie. Quindi una bolla risente progressivamente degli effetti delle spire 1, 2, 3 e viene spostata verso la spira numero 4 esattamente nel momento in cui la spira numero 1 viene nuovamente eccitata. Secondo questo schema, una bolla può trovarsi una sola volta in corrispondenza di ogni spira e può essere spostata soltanto in una direzione. Se si usa l'ordinario codice binario, la presenza di una bolla significa la cifra 1 mentre l'assenza di bolle significa 0. Le bolle possono anche essere spostate nelle due dimensioni aggiungendo un secondo sistema di spire poste ad angolo retto rispetto alle precedenti (*si veda l'illustrazione in basso a pagina 100*).

Il problema essenziale dei metodi a conduttore consiste nel fatto che è necessario connettere con i circuiti esterni una serie di conduttori di dimensioni confrontabili con quelle delle bolle, il che richiede delle operazioni assai delicate. Questo problema viene notevolmente semplificato nel sistema di guida a campo. Uno dei metodi di guida a campo consiste nel far oscillare ritmicamente il campo esterno di polarizzazione in modo tale che le bolle si espandono e si restringono alternativamente. Il movimento deriva dal fatto che la lamina è fornita di una configurazione di « trappole » energetiche molto vicine e disposte in modo asimmetrico, per esempio di una sottilissima pellicola di Permalloy costituita da piccole punte di freccia allineate approssimativamente nella medesima direzione. Quando la bolla raggiunge il suo massimo diametro, il suo centro è in corrispondenza di una di queste punte, occupando per metà la parte antistante a essa e per metà la parte posteriore. Quando il campo magnetico esterno aumenta, la bolla magnetica si contrae e il suo bordo mobile si sposta in direzione della punta. Quando il campo si abbassa nuovamente, la bolla si espande e il suo bordo si sposta verso la base della punta della freccia (*si veda l'illustrazione in alto a pagina 101*).

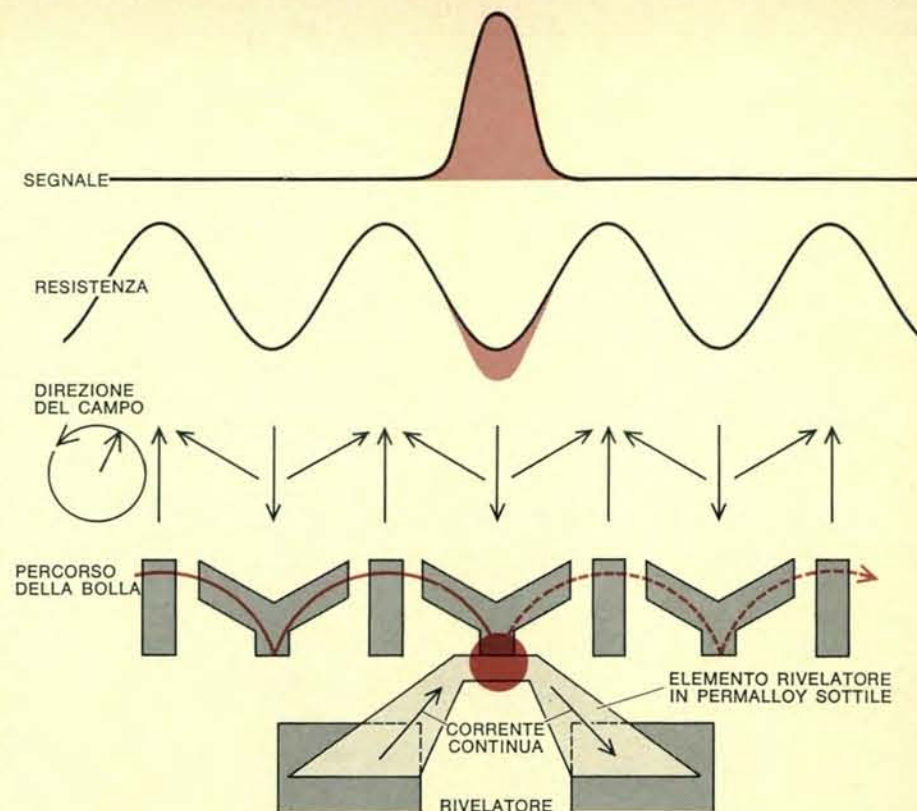
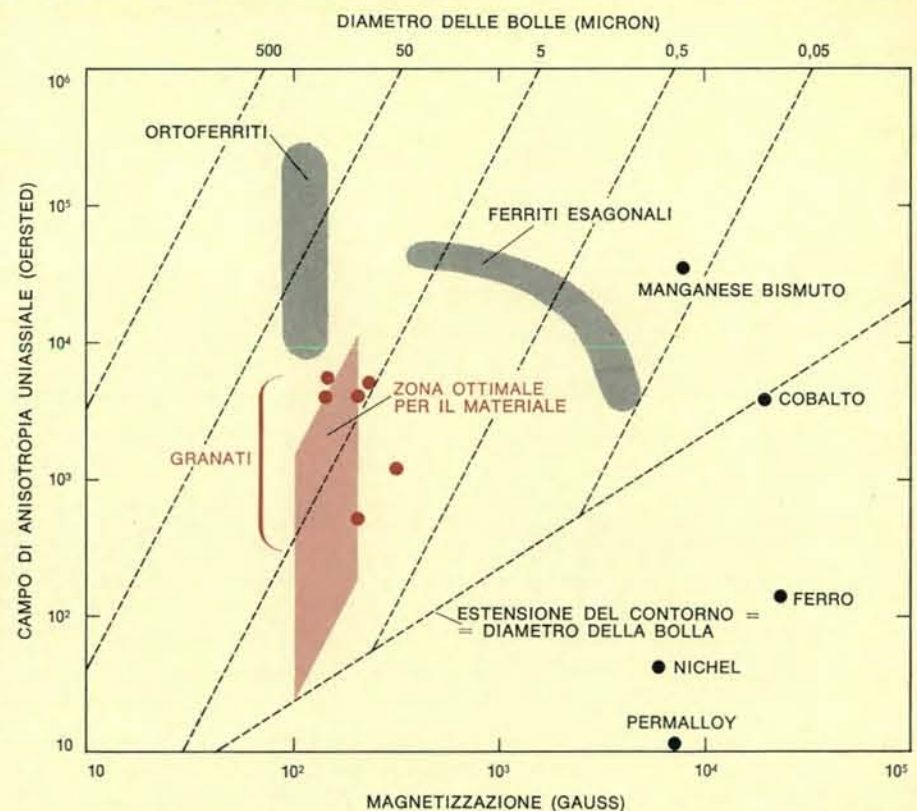
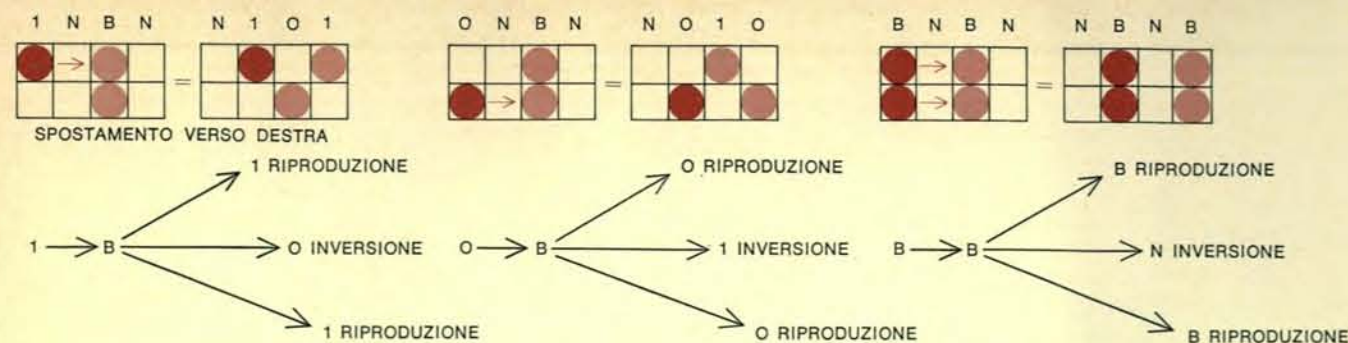
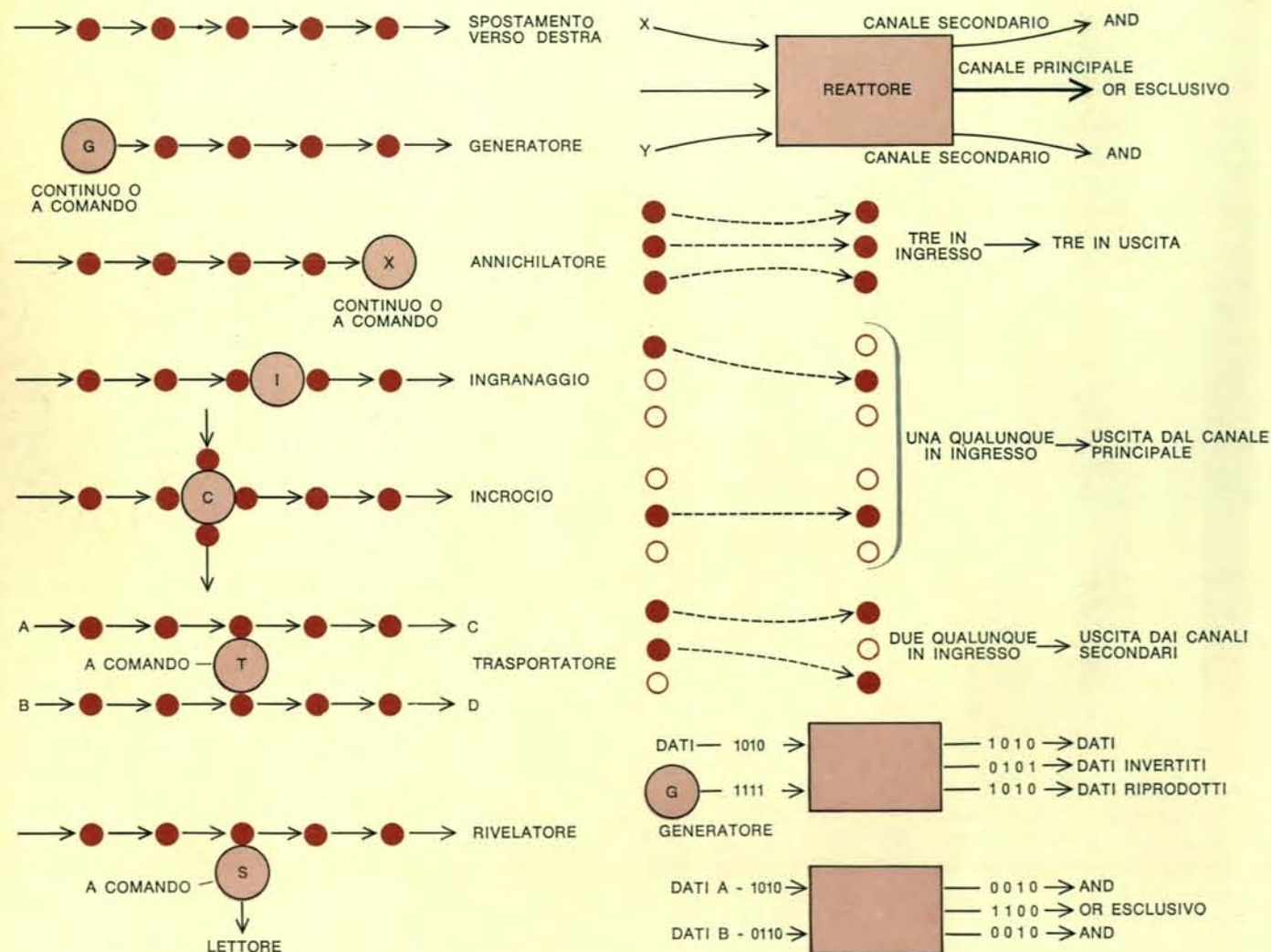
Nella figura di sinistra sono rappresentate le fondamentali operazioni logiche degli apparati a bolla. Una bolla può occupare qualunque cella di un « reticolo » (*in alto*). I vari stati di occupazione o di non occupazione di un modulo a due celle possono essere descritti mediante quattro simboli: *N*, *B*, 1, 0. Nella figura sono inoltre rappresentate cinque operazioni: inversione, riproduzione, spostamento, creazione e distruzione. È possibile dimostrare facilmente che la riproduzione implica l'esistenza di una particella che abbia un ampio raggio di interazione (*si veda l'illustrazione in alto a pagina 104*).

Un altro metodo di guida a campo, probabilmente piú vantaggioso del precedente, prevede una dimensione costante delle bolle che si spostano da punto a punto grazie a un campo magnetico rotante che agisce in collegamento con un insieme di sottili pellicole di Permalloy a forma di *T* e di lamine disposte verticalmente. Nonostante il campo ruoti, si può pensare che esso si diriga successivamente verso i quattro punti cardinali. Se in un campo di questo tipo si pone una sagoma a forma di *T*, la parte di questo *T* che, in un certo istante, è parallela al campo, si magnetizza. Quindi, quando il campo è diretto verso est risulterà magnetizzata la parte orizzontale, mentre quando il campo è diretto verso ovest risulterà magnetizzata la parte opposta. Allo stesso modo, la parte verticale risulterà magnetizzata in un senso o nell'altro a seconda che il campo sia diretto verso nord o verso sud. Se, nello stesso campo magnetico si pone una barra verticale, essa avrà una magnetizzazione diretta soltanto verso nord o verso sud; la barra stessa avrà un comportamento neutrale quando il campo sarà diretto verso est o verso ovest. Immaginiamo ora una serie di *T* posti alternati rispetto a delle barre verticali disposte in modo che la parte inferiore delle barre sia allineata con la parte orizzontale delle *T* (*si veda la seconda illustrazione dall'alto a pagina 101*). Se all'altezza della parte superiore della *T* viene introdotta una bolla dalla parte sinistra della catena e se il campo magnetico viene ruotato in senso orario, la bolla si sposterà progressivamente verso destra. Durante una rotazione completa del campo la bolla si sposta dal centro di una *T* al centro della successiva.

Ciascuno dei metodi di guida può essere progettato per produrre nuove bolle. Nello schema della guida a conduttore si può produrre una nuova bolla sistemando, su una delle spire, un conduttore a forcina che può venire eccitato separatamente. Se la forcina viene alimentata quando la bolla è in corrispondenza della spira, la bolla stessa si contrae nella sua parte centrale, si divide in due bolle che si respingono e si spostano verso due spire adiacenti (*si veda la terza illustrazione dall'alto a pagina 101*).

Nello schema di guida a campo, con un campo rotante, è possibile installare un generatore di bolle alla testa di ogni catena di *T* e di lamine: il generatore consiste in un disco di Permalloy avente una piccola sezione orizzontale. Sotto l'effetto del campo magnetico rotante si può produrre una nuova bolla tutte le volte che il campo compie una rivoluzione completa: questa nuova bolla si propaga verso destra (*si veda la quarta illustrazione dall'alto a pagina 101*). È anche molto semplice dividere in due il flusso delle bolle inserendo delle trappole o dei filtri che dirigono le bolle in eccesso verso una direzione estraendole dal flusso originario (*si veda l'illustrazione in basso a pagina 101*). La bolle cosí sottratte possono venire eliminate per mezzo di un « divoratore di bolle », una zona cioè caratterizzata da un'alta concentrazione magnetica capace di ridurre le bolle a dimensioni inferiori al loro diametro minimo di stabilità.

Ogni sistema numerico che utilizzi un sistema a bolle richiede naturalmente un metodo capace di registrare la presenza o l'assenza delle bolle stesse: il metodo di rivelazione può essere sia distruttivo sia non distruttivo. Il « divoratore di bolle », per esempio, può far parte di un sistema di rivelazione distruttivo. Le tecniche di lettura non distruttive che sono state elaborate si basano su alcuni principi fisici tra i quali ricordiamo l'induzione elettromagnetica, l'effetto Hall, la rivelazione ottica e la magnetoresistenza.

Nel metodo a induzione elettromagnetica la bolla viene utilizzata come un minuscolo dipolo magnetico che induce una debole corrente elettrica in

Questo rivelatore di bolle è basato sul principio della magnetoresistenza. Il rivelatore è una giunzione di Permalloy la cui resistenza al passaggio della corrente elettrica è una funzione sinusoidale dell'intensità del campo magnetico rotante che provoca il moto delle bolle. Quando una bolla si trova in corrispondenza della giunzione di Permalloy, la resistenza viene ridotta ulteriormente e dà quindi luogo a un segnale in uscita. Questo processo è non distruttivo: la bolla cioè continua il suo cammino.

Le superfici colorate della figura rappresentano le zone corrispondenti alle proprietà caratteristiche dei materiali a bolle. Anzitutto è fondamentale che il materiale sia « uniassiale », vale a dire che esso deve essere magnetizzabile secondo una sola direzione. Questa proprietà viene espressa per mezzo dell'intensità del campo di anisotropia uniassiale. I granati (*punti colorati*) si addensano in corrispondenza delle regioni ottimali.

Si possono facilmente ottenere delle interazioni a distanza poiché le bolle magnetiche si respingono a vicenda in modo che tra una bolla e l'altra esiste sempre un intervallo di separazione pari al diametro di una bolla. In questa illustrazione la separazione è schematicamente rappresentata soltanto in direzione orizzontale. I tre diagrammi riprodotti qui sopra descrivono le configurazioni che si ottengono quando una singola bolla, sia nello stato 1 che nello stato 0, e una coppia di bolle nello stato *B* interagiscono con le celle in uno stato *B*. In tutti e tre gli esempi lo stato iniziale viene contemporaneamente riprodotto e invertito.

Sette operazioni a flusso di bolle assolvono tutte le funzioni richieste da un sistema di memorizzazione a bolle. Quando viene usato un campo magnetico rotante per la propagazione a campo guida, le bolle si spostano costantemente in una certa direzione. Un generatore (*G*) fornisce le bolle o in modo continuo o a comando mentre l'annichilatore (*X*) può funzionare anch'esso in modo continuo o selettivo. L'ingranaggio (*I*) serve per bloccare le bolle. L'incrocio (*C*) permette che, generalmente con l'ausilio di un ingranaggio, due fasci di bolle si incrocino. Il trasportatore di bolle (*T*) trasferisce le bolle da un canale all'altro e il rivelatore (*S*) indica il passaggio di un flusso di bolle.

I flussi di bolle interagenti assolvono ulteriori funzioni logiche. Dato che le bolle interagiscono a una certa distanza, un « reattore » provvisto di tre canali di ingresso fornirà dei segnali in uscita che dipendono da quelli di ingresso. Una singola bolla uscirà sempre dal canale principale; quello centrale. Se due bolle entrano contemporaneamente nei due canali, esse usciranno dai canali secondari, ossia da quelli esterni. Secondo queste regole, i dati di ingresso rappresentati negli ultimi due diagrammi danno luogo ai dati di uscita mostrati in figura. Coloro che hanno una certa familiarità con i circuiti logici riconosceranno facilmente che questo « reattore » è sostanzialmente un addizionatore.

una spira di rivelazione posta nella testina di lettura. Nei sistemi che sfruttano l'effetto Hall, quando il campo magnetico prodotto dalla bolla agisce perpendicolarmente a una piastrina di un semiconduttore, in questa si manifesta una tensione che provoca un passaggio di corrente in senso perpendicolare alla direzione del campo magnetico. Il sistema a sensibilità ottica diretta è basato sullo stesso principio che permette di vedere i domini per mezzo di un microscopio polarizzatore: il rivelatore reagisce alle variazioni di intensità della luce provocate dal passaggio di una bolla. Il metodo a magnetoresistenza sembra il piú adatto per le microscopiche apparecchiature a stato solido che si valgono della tecnologia dei circuiti integrati: in questo metodo, la presenza di una bolla diminuisce la resistenza, e quindi la tensione, quando essa viene misurata tra i due poli di un circuito a ponte: in questo modo viene prodotto un piccolo segnale in uscita (*si veda l'illustrazione in alto a pagina 103*).

Vediamo ora come i sistemi a bolla magnetica si applicano alle esigenze dei sistemi di memorizzazione. Praticamente in tutti i calcolatori numerici le informazioni e le elaborazioni vengono fornite tramite il sistema binario. Tra le molteplici ragioni che fanno preferire il codice binario, o a due livelli, la piú importante è probabilmente rappresentata dal fatto che i migliori elaboratori e i sistemi logici posseggono appunto due stati stabili. I relé, gli interruttori, gli anelli di ferrite e i circuiti flip-flop ne sono degli esempi.

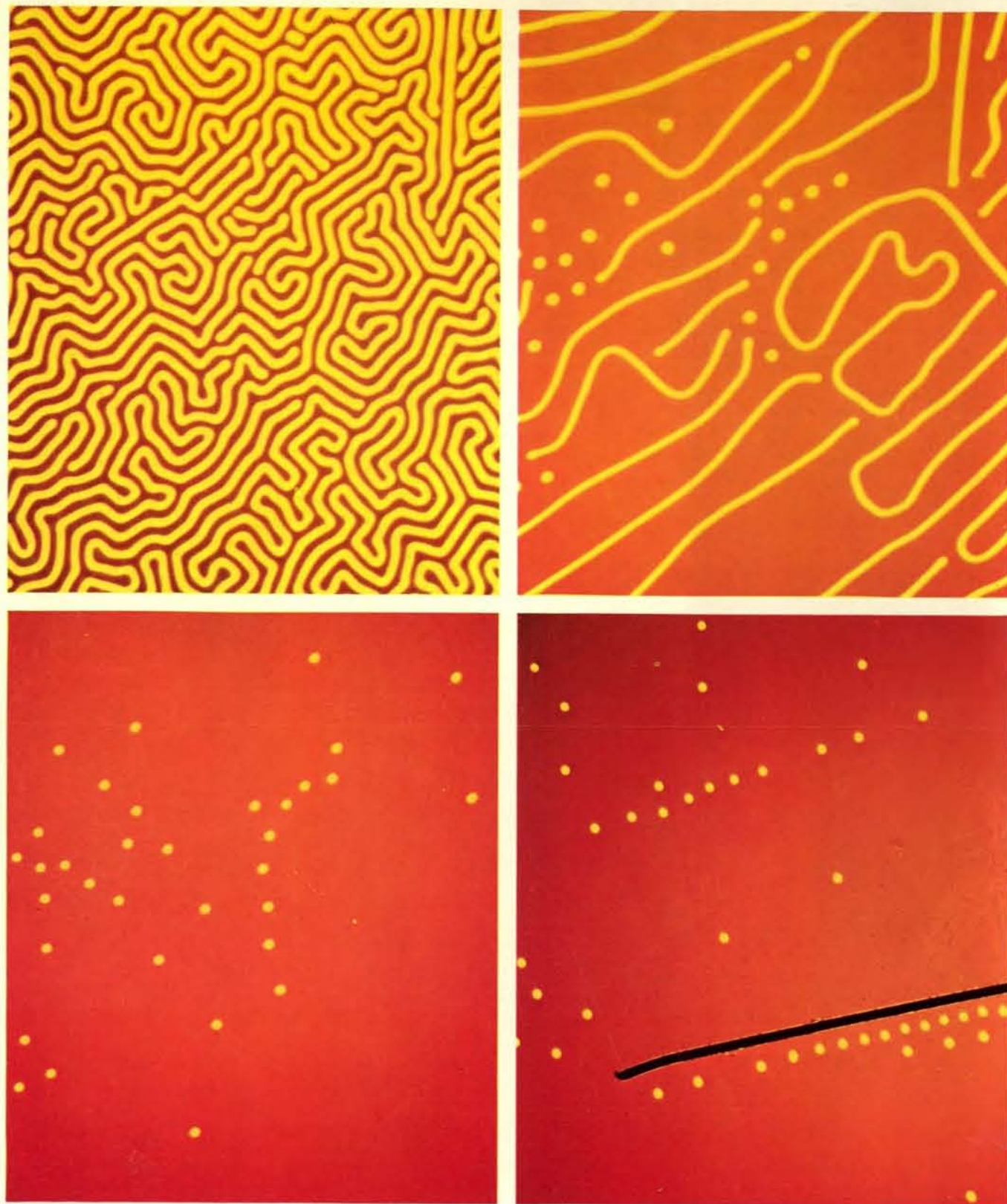
L'importanza dei sistemi a bolle per la registrazione e la logica binaria si può spiegare se si considera in che modo possono essere utilizzate a questi scopi delle particelle piú grandi, per esempio delle palle da biliardo. Nel sistema binario piú semplice, la presenza di una palla in una certa posizione ben precisa, o in una cella, può rappresentare la cifra 1, mentre la sua assenza rappresenta lo zero: pertanto, una serie di palle da biliardo intercalata da spazi vuoti potrebbe rappresentare un'informazione binaria.

Per condurre una operazione logica è comunque indispensabile che si verifichino delle interazioni: ovviamente due palle da biliardo hanno una certa correlazione dato che entrambe non possono occupare la medesima posizione contemporaneamente. Questo tipo di interazione, che definiremo a breve raggio, permette di eseguire tutte le operazioni eccettuata la riproduzione. Il processo di riproduzione, che in un sistema logico convenzionale a semiconduttore viene detta *fanout*, consiste nella duplicazione degli stati binari esistenti. Poiché generalmente i dati binari vengono utilizzati nei centri di calcolo, è essenziale che si possano riprodurre i dati per le successive rielaborazioni.

Per rendere possibile la riproduzione secondo il nostro modello a palle da biliardo, la prima modifica da fare è quella di suddividere ogni cella unitaria in due celle sovrapposte (*si veda l'illustrazione a pag. 102*), il che semplifica molte delle operazioni fondamentali. Quando nessuna delle due celle è occupata (stato *N*) si hanno delle zone disponibili perché le particelle possano essere mosse e rimesse in posizione; quando invece entrambe le celle sono occupate (stato *B*) si ha una riserva di particelle. Gli stati corrispondenti a una configurazione 1 e 0 sono simmetrici e vengono definiti arbitrariamente: una particella nella cella superiore significa 1 mentre una nella cella inferiore significa zero; è evidente che, con una semplice inversione delle celle, ogni 1 si trasforma in uno 0 e viceversa.

Si ottiene la riproduzione introducendo il concetto di interazione a lungo raggio. Ammettiamo che, tutte le volte che una particella penetra in una cella adiacente a un'altra cella occupata, essa respinga verso destra la particella occupante in modo tale che esista sempre almeno una cella vuota tra due particelle. In altri termini, ogni particella possiede una sfera di influenza pari al doppio del suo diametro. Dobbiamo comunque supporre, almeno per quanto riguarda i diagrammi, che la sfera di influenza sia limitata in due dimensioni su di un piano orizzontale; senza questa convenzione, i nostri diagrammi dovrebbero tradurre il fatto che le due celle verticali del modulo di base debbono includere anche una cella vuota.

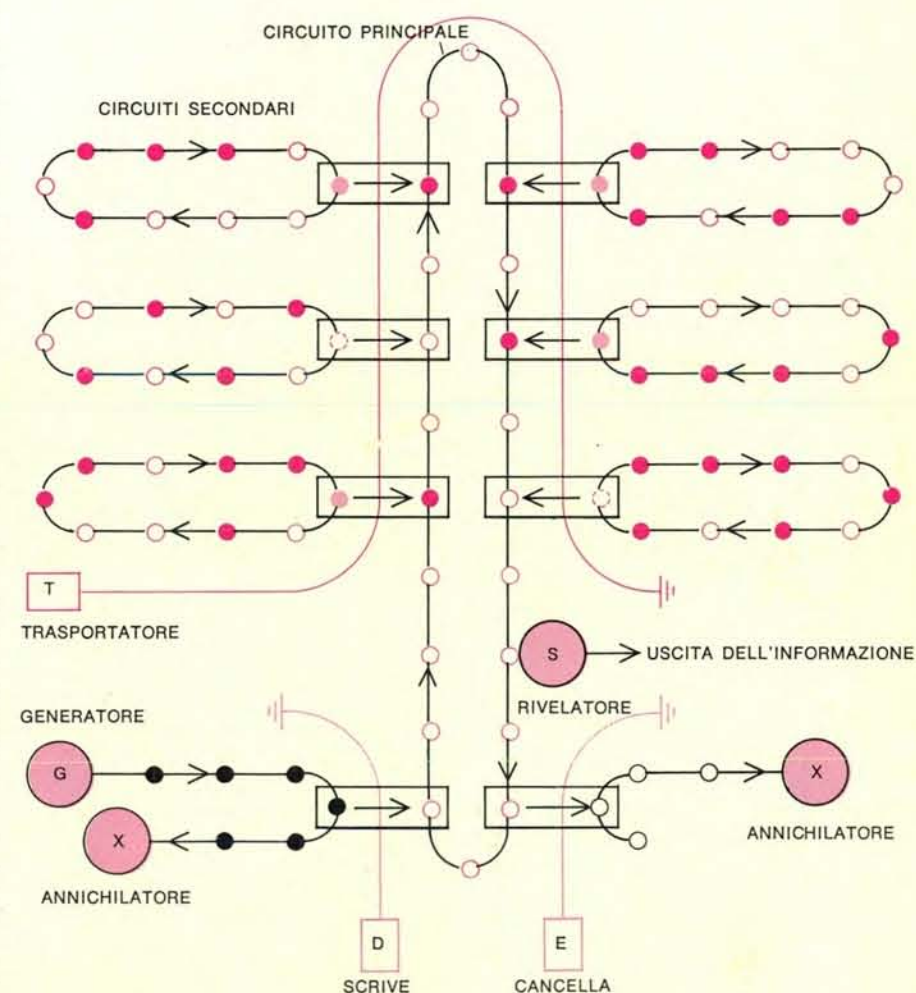
Dato che le bolle magnetiche dei circuiti che abbiamo descritto piú sopra posseggono una sfera di influenza pari al doppio del loro diametro, a esse possiamo applicare perfettamente il nostro modello a palle da biliardo. Ora, come si può osservare, se una bolla nello stato 1 interagisce con uno stato *B*, gli stati che ne risultano sono 1, 0, 1. In altre parole, questo tipo di interazione provoca una riproduzione e una inversione (*si veda l'illustrazione in alto nella pagina a fronte*). Circuiti capaci di compiere questo tipo di operazioni logiche e altre simili sono già stati costruiti e collaudati.

L'aspetto piú interessante della nuova tecnologia delle bolle magnetiche consiste nella possibilità di riunire in un solo apparato microscopico sia l'ac-

Questa serie di microfotografie eseguite nei Laboratori della Bell Telephone mostra l'apparire di bolle magnetiche in una sottile lamina di un granato magnetico. A seconda che i campi magnetici interni al cristallo siano diretti verso l'alto o verso il basso, i domini magnetici del campione provocano una rotazione della luce polarizzata secondo differenti direzioni. Mediante opportuni filtri polarizzatori i domini orientati nella medesima direzione possono risultare chiari o scuri. In assenza di campi magnetici esterni (*in alto a sinistra*) i domini formano delle figure ondulate in cui i domini aventi magnetizzazione opposta occupano la stessa superficie. (In questa fotografia, si nota una apparente disuguaglianza dovuta a un difetto di esposizione). Quando viene applicato un campo magnetico perpendicolare al campione (*in alto a destra*), i domini che hanno magnetizzazione opposta si contraggono e, in alcuni casi, danno luogo a bolle. In realtà, nonostante questi domini circolari vengano chiamati bolle, essi sono dei cilindri tronchi visti da un'estremità. Se si aumenta ulteriormente il valore del campo magnetico esterno (*in basso a sinistra*), anche i rimanenti domini si trasformano in bolle. È possibile spostare liberamente le bolle mediante un filamento magnetico la cui magnetizzazione sia polarizzata per mezzo di un campo magnetico esterno (*in basso a destra*). Dato che tra le bolle esiste una forza repulsiva, esse tendono a mantenere tra loro una certa distanza minima: nonostante ciò esse possono essere raggruppate sino a raggiungere una densità dell'ordine del milione di bolle al centimetro quadrato.

cumulazione di dati sia la logica. Nei sistemi di memoria piú semplici, le bolle possono essere introdotte, spostate e rivelate con un registratore mobile a spira anulare. Per quanto riguarda il modo in cui è possibile attuare un registratore di questo tipo, i pareri sono discordi. Il sistema che noi abbiamo trovato piú conveniente è quello di guida a campo in cui si fa ruotare un campo piano con una struttura di *T* e di barre o con una struttura di forma analoga. Sono stati collaudati dei registri dell'ordine di 10 000 passi, in una piastrina delle dimensioni di 2 millimetri di lato (*si vedano le illustrazioni a pagina 96*).

Lo schema dei progetti dei sistemi di memorie di massa consiste in una serie di registri mobili a spire anulari collegati tra loro. I dati, introdotti nelle cosiddette spire secondarie, vengono trasferiti per la lettura e per le eventuali rielaborazioni in un'unica spira principale. La parte piú importante dello schema consiste nell'operazione durante la quale le bolle vengono trasferite dalle spire secondarie a quella principale e viceversa. Lo spostamento deve avvenire in perfetto accordo col ciclo del campo magnetico rotante. Per cancellare una bolla basta spostarla in una spira secondaria provvista di un « divoratore di bolle ». Per introdurre i dati si trasferiscono le bolle in modo selettivo da un « serbatoio di bolle » che in effetti non è altro che una spira secondaria avente da un lato un generatore di bolle e dall'altro un divoratore. Il campo ruota in continuazione per la durata di tutte le operazioni. Come prototipo di memoria di massa a bolle abbiamo progettato e costruito una specie di « elenco telefonico » che può contenere da 50 a 100 numeri telefonici di uso piú comune (*si veda l'illustrazione in questa pagina*).

Il metodo piú promettente per produrre dei materiali con una struttura a bolle magnetiche è quello di farli sviluppare sotto forma di pellicole epitassiali di spessore compreso fra i tre e i dieci millimicron su un substrato monocristallino non magnetico: una pellicola epitassiale di questo tipo riproduce la struttura del reticolo cristallino del substrato. Per produrre queste pellicole sono stati sperimentati con successo sia dei sistemi a fase liquida sia a vapore chimico. In piastrine capaci di contenere fino a quindici memorie, ognuna delle quali dotata di una capacità di 4000 bit, sono state ottenute delle densità di difetti molto basse, dell'ordine di 10 per centimetro quadrato. Sono stati inoltre costruiti dei registri sperimentali con delle densità di memoria attorno al mezzo milione di bit al centimetro quadrato e con delle velocità di elaborazione superiori a un milione di bit al secondo. Pensiamo che sarà possibile sviluppare i nostri circuiti di propagazione e di rivelazione direttamente nelle piastrine epitassiali.

I progressi ottenuti nel campo dei materiali a bolle come nel campo dei collegamenti elettrici rendono possibile oggi la costruzione di registri a bolla a 20 000 fasi in una superficie di due millimetri quadrati che corrisponde alla superficie che era prima necessaria per un registro a 20 fasi. Sarebbe possibile produrre una memoria a bolla avente una capacità di accumulazione di 10^7 bit in un volume di 50 centimetri cubi usando una serie di strati sovrapposti. Per far muovere i dati a una velocità di 200 000 bit al secondo è sufficiente una potenza di soli 10 watt. Dato che i domini delle bolle magnetiche vengono spostati da punto a punto da sistemi di elementi magnetici che vengono alimentati da un campo magnetico rotante generato all'esterno, è sufficiente un numero molto limitato di collegamenti con l'esterno. Possiamo senz'altro prevedere che i sistemi a bolla magnetica permetteranno di ottenere, a un costo molto basso, elevate capacità di accumulazione di informazioni con affidabilità molto alta.

L'intero complesso di una memoria a bolle può essere organizzato secondo una struttura tale che una certa quantità di spire minori fornisca i dati e li trasferisca a comando in un circuito maggiore di lettura. I dati sono stati trasferiti da sei circuiti secondari nel circuito principale (che in precedenza era stato svuotato) e sono qui rappresentati in colore chiaro. Di conseguenza, nelle spire secondarie si manifestano dei vuoti che possono essere sostituiti o con i dati precedenti o con dei nuovi dati. I nuovi dati vengono inseriti nella memoria trasportando delle bolle selezionate dalla spira serbatoio di bolle (*G-X*) secondo le informazioni impartite dal comando « scrittura » (*D*). I dati vengono cancellati dalla memoria per mezzo di un operatore (*E*) e di un trasferitore di bolle dal circuito principale all'annichilatore (*X*) posto in basso a destra. I complessi dotati di circuiti principale e secondario hanno una capacità di 20 000 bolle ciascuno e, quando vengano disposti in modo intercalato a struttura rettangolare, permettono un accesso casuale al flusso di dati in una memoria di 15 milioni di bit. Mentre *T* ed *E* vengono collegati orizzontalmente, *D* e *S* lo sono verticalmente.

GIOCHI MATEMATICI

di Martin Gardner

Il paradosso del dado non transitivo e il principio di indifferenza

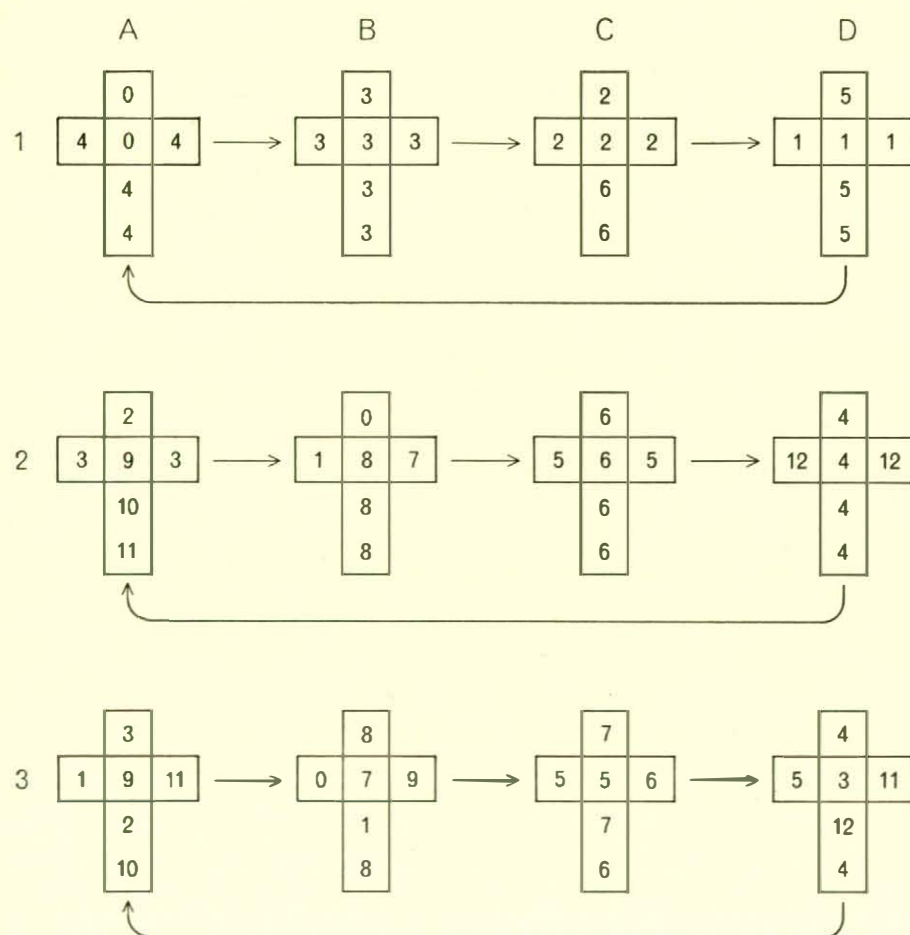
Nella teoria della probabilità abbondano casi paradossali capaci di trarre in inganno il senso comune e di far cadere in errore lo sprovveduto. Questo mese tratteremo un nuovo strabiliante caso paradossale che riguarda la proprietà delle relazioni detta transitività e un gruppo di paradossi che trovano la loro origine in una poco accorta applicazione di quello che viene detto principio di indifferenza.

La transitività è una proprietà di una relazione binaria per cui, se la relazione sussiste fra *A* e *B*, e anche fra *B* e *C*, allora deve sussistere necessariamente anche tra *A* e *C*. Un esempio molto comune ci è dato dalla relazione « piú pesante di ». Se *A* è piú pesante di *B*, e *B* è piú pesante di *C*, allora anche *A* è piú pesante di *C*. I tre insiemi, formati ciascuno dai quattro dadi la cui superficie è stata « sviluppata » nella figura qui sotto, sono stati ideati da Bradley Efron, uno studioso di statistica della Stanford University per illustrare chiaramente alcune recenti scoperte relative a una generica classe di conclusioni paradossali che riguardano fenomeni probabilistici non soggetti alla transitività. Per ciascuno di questi insiemi si possono fare giochi di scommesse cosí poco intuitivi che anche per giocatori esperti risulterà impossibile rendersi conto della situazione anche dopo averla analizzata molto attentamente.

I dadi non transitivi.

I quattro dadi del primo insieme della figura sono numerati nel piú semplice modo che dà la possibilità di vincere col massimo vantaggio. Fate scegliere a qualcuno un dado di questo primo insieme. Poi anche voi sceglietene uno fra i tre rimanenti. Gettati entrambi i dadi vincerà chi otterrà il numero piú alto. Sembrerebbe che, se il vostro antagonista ha la possibilità di scegliere per primo a ogni mano, il gioco dovrebbe essergli favorevole o al massimo terminare in parità. Se almeno due dadi hanno la stessa probabilità di vittoria, il gioco è equo, perché se l'altro prende uno dei due dadi, voi potete prendere il rimanente; se invece un dado è migliore degli altri tre, chi sceglie per primo può sempre prendere proprio quel dado e vincere piú della metà delle mani. In realtà questo discorso risulta completamente sbagliato. L'incredibile verità è che, indipendentemente da ciò che prende l'altro giocatore, voi potete sempre scegliere un dado che ha una probabilità di vincere pari a 2/3, cioè ogni tre mani due vi saranno favorevoli!

Il paradosso (che è tale in quanto va contro il senso comune) sorge dall'errore iniziale di considerare la relazione « aver maggiore probabilità di vincere » una relazione transitiva se applicata a coppie di dadi. Ma ciò non accade per i dadi di ciascuno di questi tre insiemi. In ogni insieme la relazione « aver maggiore probabilità di » è rappresentata da una freccia orientata verso il dado perdente. Il dado *A* batte *B*, *B* batte *C*, *C* batte *D*, e *D* batte *A*! Nel primo insieme la probabilità di vincere col dado indicato di ogni coppia di dadi è 2/3. Questo valore si può controllare facilmente facendo il rapporto tra le 36 uscite possibili e i 24 casi in cui un dado contiene un numero maggiore dell'altro.

I dadi degli altri due insiemi, sempre progettati da Efron, godono della stessa proprietà di non-transitività, però contengono meno numeri ripetuti per renderne piú difficile l'analisi. Anche nel secondo insieme, presi due dadi nell'ordine indicato dalle frecce, la probabilità che il primo batta il secondo è 2/3. Nel terzo insieme, poiché sono possibili casi di parità, si deve convenire che in tali casi si rilancino i dadi fino a che uno dei due non superi l'altro. Se si gioca attenendosi a queste norme, la probabilità di vincita per ogni coppia di dadi del terzo insieme è 22/34, cioè 0,647.

È stato dimostrato inoltre, afferma Efron, che 2/3 è il massimo vantaggio raggiungibile in un insieme di quattro dadi. Per tre insiemi di numeri, il vantaggio massimo è 0,618, ma è impossibile ottenerlo con i dadi poiché gli elementi degli insiemi devono contenere piú di sei numeri. Se si impiegano piú di quattro insiemi (con numeri scelti a caso all'interno di ogni insieme) il possibile vantaggio si avvicina sempre piú al limite 3/4 man mano che cresce il numero degli insiemi.

Esiste un principio fondamentale nel calcolo delle probabilità di eventi del tipo considerato (uscita di un numero al lancio di dadi) che risale al XVIII secolo, agli inizi cioè della teoria classica della probabilità. In un primo tempo era stato chiamato « principio di ragione insufficiente »; ora è conosciuto come « principio di indifferenza »; questa denominazione piú semplice è stata usata da John Maynard Keynes nel suo libro *A Treatise on Probability*. (Keynes è forse piú noto come economista, sebbene il suo libro sulla probabilità sia ormai considerato un classico. Esso ebbe una decisiva influenza sulla logica induttiva dell'ultimo Carnap). Il principio di solito si enuncia cosí: se non vi sono ragioni per credere che uno di *n* eventi che si escludono a vicenda possa accadere con maggiore probabilità degli altri, allora a ognuno di questi eventi si assegna la probabilità $1/n$.

Per esempio: si osserva un dado con molta attenzione e non si trova nulla che possa favorire una faccia piú di un'altra, cosa che potrebbe accadere se ci fosse da qualche parte un peso nascosto, oppure se la forma del dado non fosse perfettamente cubica a causa di smussature presenti su alcuni bordi, oppure per una maggiore scivolosità di alcuni lati, ecc. Si può allora affermare che il dado può presentarsi in sei posizioni diverse tutte egualmente probabili; pertanto si assegna a ciascuna di esse probabilità 1/6. Se si lancia una moneta, oppure, dati due cubetti di zucchero, si scommette alla maniera messicana, su quale dei due una mosca si poserà prima, non essendoci motivo per pensare che un caso sia piú favorevole di un altro, si è portati ad attribuire a ciascuno dei due possibili eventi una probabilità di 1/2. Non è necessario fare statistiche o prove sperimentali per nessuno degli esempi fatti: le probabilità vengono assegnate a priori, in base a considerazioni sulla simmetria della struttura e delle forze che possono influire sull'evento. Il dado è un solido regolare: la probabilità che la moneta resti in bilico appoggiata sul bordo è in pratica nulla; non esiste alcun motivo logico per credere che una mosca sceglierà una zolletta di zucchero piuttosto che un'altra, ecc. In ultima istanza, ovviamente, la nostra analisi è fondata su fattori di tipo sperimentale, poiché solo l'esperienza può far vedere, per esempio, che la presenza di un peso su una faccia di un dado può influire sull'uscita di un numero dispari, mentre dipingendo la stessa faccia di rosso (e tutte le altre di blu) non si avrebbe alcun effetto particolare.

Una qualche forma del principio di indifferenza è indispensabile nella teoria della probabilità; bisogna però caratterizzarlo con cura e applicarlo con estrema cautela se non si vuole incorrere in errori. In molti casi la difficoltà maggiore, che spesso trae in inganno, stà nel decidere quali sono gli eventi egualmente probabili. Si supponga, per esempio, dopo aver mescolato quattro carte – due rosse e due nere – di disporre queste carte in fila e coperte. Si sollevino poi due carte a caso. Qual è la probabilità che le due carte siano dello stesso colore?

Qualcuno potrebbe fare questo ragionamento: « Ci sono tre casi egualmente probabili. Le carte possono essere o entrambe nere, o entrambe
lore diverso. Allora in due casi le carte sono dello stesso colore, e pertanto la probabilità che l'evento richiesto si verifichi è 2/3. »

Qualcun altro potrebbe obiettare: « Questo non è vero: ci sono *quattro* casi egualmente probabili; infatti le due carte possono essere tutte e due nere, tutte e due rosse, la prima nera e la seconda rossa oppure la prima rossa e la seconda nera. Piú semplicemente, o le carte sono dello stesso colore o non lo sono. E in qualunque modo risultino accoppiate, la probabilità che siano dello stesso colore è 1/2 ».

In realtà questi due ragionamenti sono entrambi sbagliati. (La vera probabilità verrà data il prossimo mese. Provi il lettore a calcolarla da solo.) Gli errori commessi nell'uno e nell'altro derivano dall'aver considerato egualmente probabili eventi che non lo sono. Esistono tuttavia paradossi addirittura sconcertanti – vere e proprie fallacie – in cui a prima vista sembra possibile un'applicazione del principio di indifferenza, e d'altra parte applicandolo si giunge immediatamente a una contraddizione logica. Casi di questo tipo si hanno quando non esiste una ragione valida per considerare *n* eventi egual-

mente probabili e l'ipotesi di equiprobabilità è fondata interamente, o quasi, sull'ignoranza.

Per esempio, qualcuno ci dice: «Nella stanza accanto c'è un cubo la cui dimensione è stata scelta a caso. Si sa che lo spigolo del cubo è compreso tra un metro e tre metri.» Come si potrebbe stimare la probabilit che lo spigolo del cubo sia compreso tra 1 e 2 metri in rapporto alla probabilità che sia compreso tra 2 e 3 metri? Non sarebbe logico, in mancanza di altre informazioni, rifarsi al principio di indifferenza e considerare ciascuna probabilità uguale a 1/2?

Anche in questo caso tuttavia, applicando il principio di indifferenza si sbaglierebbe. Infatti, se lo spigolo del cubo varia tra 1 e 2 metri, il suo volume varia tra 1^3, cioè un metro cubo, e 2^3, cioè otto metri cubi. Quando però lo spigolo va dai 2 ai 3 metri, il volume subisce un aumento compreso tra 2^3 (otto metri cubi) e 3^3 (ventisette metri cubi) e l'intervallo di variazione è quasi il triplo dell'altro. Se il principio di indifferenza è applicabile ai due intervalli in cui variano gli spigoli esso non lo è piú in relazione ai corrispondenti intervalli dei volumi. Non si era chiarito all'inizio quale « dimensione » del cubo era stata scelta a caso, e poiché il termine « dimensione » è piuttosto vago (potrebbe riferirsi sia allo spigolo del cubo, sia al suo volume) non si hanno dati su cui fondare le proprie supposizioni. Il principio di indifferenza si potrebbe applicare se fosse stato scelto a caso lo spigolo del cubo. Esso sarebbe applicabile anche se si fosse detto che era stato scelto a caso il volume del cubo, ma naturalmente in questo caso la probabilità 1/2 andrebbe attribuita ai due intervalli di volume che vanno da 1 a 14 m^3 e da 14 a 27 m^3 e di conseguenza ai corrispondenti intervalli di variazione dei lati. Ne deriva che se il principio si applica allo spigolo, non si può applicare al volume senza cadere in contraddizione, e viceversa. E dal momento che non si sa con esattezza di quale dimensione si tratti, in nessun caso ha senso applicare il principio di indifferenza.

Carnap, per dimostrare l'uso indiscriminato che di questo principio si fa nel libro *Theory of Probability* di Harold Jeffreys, propone il seguente esempio. Si sa che in un'urna ci sono delle palline colorate di blu, di rosso o di giallo, ma non si sa quante palline ci siano per ciascun colore. Qual è la probabilità che la prima pallina estratta dall'urna sia blu? Applicando il principio di indifferenza si potrebbe subito dire 1/2. Naturalmente è 1/2 anche la probabilità che la pallina non sia blu. Se non è blu, deve per forza essere rossa o gialla, e poiché non si conosce il numero delle palline rosse e gialle, entrambi i colori risultano equiprobabili. Pertanto applicando ancora una volta il principio si può dire che verrà estratta una pallina rossa con probabilità 1/4. D'altra parte, se inizialmente si vuole la probabilità che la prima pallina estratta sia rossa, secondo il precedente ragionamento bisogna attribuire a tale probabilità il valore 1/2, mentre diventerebbe 1/4 la probabilità che la pallina sia blu; il che è in contraddizione con le conclusioni appena raggiunte.

Con procedimenti analoghi si possono dimostrare facilmente le cose piú strane come per esempio che su Marte esistono forme di vita. Che probabilità c'è che su Marte si trovi una semplice forma di vita di tipo vegetale? Poiché esistono argomenti egualmente convincenti che appoggiano entrambe le tesi (che la vita possa o non possa esistere) si può rispondere che la probabilità è 1/2. E che probabilità c'è di trovare su Marte anche una semplice forma di vita animale? Ancora 1/2. A questo punto si è costretti ad affermare che la probabilità che su Marte esista « o la vita vegetale o la vita animale » è data da 1/2 + + 1/2 = 1, il che equivale a certezza; ma questo è assurdo. Anche il filosofo Charles Sanders Pierce, screditò il principio di indifferenza per mezzo di argomentazioni di questo genere, dimostrando che i capelli degli abitanti di Saturno devono essere di due colori diversi. Molti altri esempi di falsi ragionamenti simili a questi si possono trovare nel quarto capitolo del libro di Keynes. È facile inventarne anche altri.

Nella storia della metafisica uno dei casi piú famosi in cui è stato usato in modo sbagliato questo principio è certamente quello di Blaise Pascal, pioniere della teoria della probabilità, in un famoso ragionamento noto sotto il nome di « scommessa di Pascal ». Vale la pena di riportare qualche brano della trattazione originale (*Pensieri*, Pensiero n. 233) anche se quest'ultima risulta alquanto prolissa.

« Dio esiste oppure no. Quale delle due tesi dovremo sostenere? Razionalmente non si può decidere nulla. Siamo ai capi opposti di un infinito abisso. Alle estremità di questa enorme distanza si sta facendo un gioco in cui tirando una monetina può venire fuori testa o croce. Che cosa scommetterai? Non c'è nessuna ragione valida per puntare su una affermazione piuttosto che sull'altra, non puoi razionalmente decidere in favore di una delle due....

« È vero; ma sei costretto a scommettere su qualcosa... Cosa sceglierai? Soppesiamo bene quello che c'è da guadagnare o da perdere nel scegliere lo evento "testa" e cioè "Dio esiste". Se vinci, vinci tutto. Se perdi, non perdi nulla. Allora scommetti subito e senza esitare che Dio esiste. »

Lord Byron, in una lettera ripropone in effetti il pensiero di Pascal: « Indubbiamente, chi possiede una fede incrollabile nel Vangelo ha un notevole vantaggio su tutti gli altri, per il semplice motivo che, se è vero ciò in cui crede, avrà un giorno una ricompensa; e se questo giorno non ci sarà, gli accadrà tutt'al piú di dormire il suo sonno eterno accanto a un miscredente, avendo però avuto in vita il conforto della fede, e nessuna delusione in seguito, poiché (nel peggiore dei casi) dal nulla non può nascere nulla, neppure di doloroso. » Brani di questo tipo si possono trovare anche in molti scritti apologetici religiosi contemporanei.

Pascal non fu il primo a sostenere che la scelta migliore in questa scommessa era la fede nell'ortodossia cristiana. Una argomentazione analoga era già stata usata dal sacerdote africano Arnobio il Vecchio nel IV secolo, e altre ancora, ma non riguardanti il Cristianesimo, risalgono a Platone. Non è questo comunque il luogo per addentrarsi nella curiosa storia delle difese e delle critiche fatte a questa scommessa. Mi accontenterò solo di ricordare che Denis Diderot osservò che questa stessa scommessa può essere ugualmente efficace se applicata ad altre fedi, per esempio all'islamica. Dal punto di vista matematico l'aspetto interessante di tutto ciò è che Pascal paragona il risultato della sua scommessa al verificarsi di un certo evento nel lancio di una moneta. In altre parole, egli applica esplicitamente il principio di indifferenza a una situazione in cui applicarlo è matematicamente privo di senso.

Tra le moderne riformulazioni del principio di Pascal, la piú arguta è quella contenuta nel famoso saggio di William James *The Will to Believe* in cui si sostiene che, considerato come gioco d'azzardo, il teismo filosofico è una scelta migliore dell'ateismo. In forma ancor piú blanda tale argomentazione è usata ogni tanto dagli umanisti per sostenere la tesi ottimista contro quella pessimista in un tempo in cui l'estinzione della razza umana in un prossimo futuro sembra altrettanto probabile che la sua sopravvivenza.

« Finché il mondo ha una sola possibilità di superare tutte le sue difficoltà », dice il narratore della poco nota novella di H.G. Wells *Apropos of Dolores*, « ritengo che un uomo ragionevole si debba comportare come se fosse sicuro di ciò. Se alla fine il vostro ottimismo non sarà giustificato, almeno sarete stati allegri ».

Il mese scorso avevamo chiesto ai lettori di studiare le proprietà cicliche del periodo di 1/13. Questo periodo, 076 923, non è un vero ciclico nel senso definito il mese scorso; è però ciclico in questo secondo duplice senso. Se viene moltiplicato per qualsiasi numero da 1 a 12, metà dei prodotti sono le sei permutazioni cicliche di 076 923, mentre l'altra metà sono le sei permutazioni cicliche di 153 864. Si noti che entrambi questi numeri (come il piú piccolo ciclico di ordine 1, 142 857) possono essere spezzati in due parti che, sommate assieme, danno 999. Possono anche essere divisi in tre parti che, sommate assieme danno 99 (07 + 69 + + 23 = 99; 15 + 38 + 46 = 99; 14 + + 28 + 57 = 99).

Quando un numero di *n* cifre, moltiplicato per tutti i numeri da 1 a $2n$, porta a dei prodotti che sono tutti permutazioni cicliche di due numeri di *n* cifre, viene detto ciclico di ordine 2. Trascurando il caso banale generato da 1/3, il piú basso numero primo che genera un ciclico di ordine 2 è 13. Gli altri primi inferiori a 100 che generano ciclici di ordine 2 sono 31, 43, 67, 71, 83 e 89.

I numeri ciclici possono essere di ordine qualsiasi. Il piú piccolo primo che genera un ciclico di ordine 3 è 103. Il periodo di 1/103, moltiplicato per ogni intero da 1 a 102 dà dei prodotti che cadono in tre insiemi, ciascuno dei quali contiene 34 permutazioni cicliche di numeri di 34 cifre. Il piú piccolo primo che genera un ciclico di ordine 4 è 53. In generale, se il reciproco di un primo *p* ha un periodo decimale ripetuto con una lunghezza (numero di cifre) pari a $(p - 1)/n$, il periodo è un ciclico di ordine *n*. Per esempio, 1/37 genera il periodo di tre cifre 027; poiché 36/3 = = 12, il periodo è un ciclico di ordine 12. I piú bassi numeri primi che generano ciclici di ordine da 5 a 10 sono rispettivamente 11, 79, 211, 41, 73 e 281. Si noti che i 10 prodotti del ciclico di ordine 5 (09, il periodo di 1/11) sono i primi 10 multipli di 9.

Come hanno notato molti lettori, il destino della svastica nel gioco della « Vita » è diverso: non si trasforma in un semaforo alla terza mossa bensí svanisce alla sesta. Anche i destini degli 11 pentomini possibili sono errati: cinque (e non sei) svaniscono rapidamente e quattro (non tre) diventano semafori.